



TAMPEREEN TEKNILLINEN YLIOPISTO  
TAMPERE UNIVERSITY OF TECHNOLOGY

TOMMI HIRVONEN

TUTKIMUSDATAN HALLINNAN KEHITTÄMINEN JA DATAN  
LAADUN ARVIOINTI ASiantuntijayrityksessä

Diplomityö

Tarkastaja: professori Samuli Pekkola  
Tarkastaja ja aihe hyväksytty  
Talouden ja rakentamisen tiedekunta-  
neuvoston kokouksessa 4.11.2015

## TIIVISTELMÄ

**TOMMI HIRVONEN:** Tutkimusdatan hallinnan kehittäminen ja datan laadun arviointi asiantuntijayrityksessä  
Tampereen teknillinen yliopisto  
Diplomityö, 90 sivua, 8 liitesivua  
Kesäkuu 2016  
Tietojohtamisen diplomi-insinöörin tutkinto-ohjelma  
Pääaine: Tietohallinto ja -järjestelmät  
Tarkastaja: professori Samuli Pekkola

**Avainsanat:** datan hallinta, datan laadun arviointi, datan hallinnointi

Yrityksen käyttämän datan laatu määrittelee, kuinka hyviä kyseisen yrityksen tekemät raportit ja päätökset voivat olla. Datan laatu on merkittävä tekijä yritysten liiketoiminnalle, mutta datan valtava määrä nykyaikaisissa yrityksissä asettaa haasteita datan hallinnalle. Tehokkaalla datan hallinnalla voidaan osaltaan varmistaa datan hyvä laatu niin lyhyellä kuin pitkällä aikavälillä. Datan hallinnassa ei ole kyse kuitenkaan pelkästä teknisestä ongelmasta, vaan siihen liittyvät oleellisesti teknologian lisäksi myös ihmiset ja prosessit. Nämä organisatoriset näkökulmat on olennaista ottaa huomioon datan hallinnassa.

Tässä työssä selvitettiin, miten yrityksen datan hallintaa voidaan kehittää ja datan laatua arvioida. Työn kohdeyrityksenä toimi suomalainen keskikokoinen asiantuntijayritys ja tarkasteltavana datana kyseisen yrityksen tutkimusdata. Kirjallisuuskatsauksen avulla esiteltiin datan laatuun, sen arviointiin, datan hallinnointiin sekä datan hallinnan ja laadun kehittämiseen liittyvää teoriaa. Työssä arvioitiin kohdeyrityksen tutkimusdatan laatua käyttäen subjektiivista ja kvalitatiivista haastattelumenetelmää. Lisäksi kohdeyrityksen datan hallinnan ongelmia kartoitettiin haastattelemalla, ja näiden ongelmien pohjalta yritykselle määriteltiin datan hallinnan kypsyyden nyky- ja tavoitetaso. Lopuksi kohdeyritykselle luotiin kehitysehdotukset, jotka auttavat yritystä kehittämään datan hallintaa kohti tavoitetasoa.

Datan laadun ulottuvuudet ovat datan piirteitä, joita voidaan mitata. Näiden ulottuvuuksien toteutumista voidaan arvioida erilaisilla vakiintuneilla menetelmillä. Menetelmät voivat hyödyntää objektiivisia ja/tai subjektiivisia mittareita laadun arvioinnissa. Tässä työssä esiteltiin kolme datan laadun arvioinnin menetelmää, joita yritykset voivat hyödyntää omassa laadun arvioinnissaan.

Yritykset voivat kehittää datan hallintaansa muun muassa hyödyntämällä datan hallinnoinnin viitekehyksiä. Datan hallinnointi on roolien, vastuiden, käytäntöjen ja prosessien määrittelyä niin, että yritys voi varmistaa datan tehokkaan hallinnan. Datan isännöinti taas varmistaa sen, että hallinnoinnin strategiset linjaukset muutetaan taktiseksi suunnitelmaksi ja jalkautetaan yrityksen jokapäiväisessä toiminnassa. Datan hallinnan kehittämisen hankkeissa on olennaista tunnistaa kriittiset onnistumistekijät, jotka vaikuttavat hankkeen onnistumiseen. Kirjallisuus tarjoaa myös datan hallinnan parhaita käytäntöjä, joilla yritykset voivat kehittää toimintaansa. Datan hallinnan kehittämisen lisäksi turvattava sen vaatimat resurssit. Ne mahdollistavat kehityksen, mutta ne on myös osattava hyödyntää oikein. Olennaista datan hallinnan kehittämisen hankkeissa on liiketoiminnan ja tietohallinnon yhteistoiminta. Tietohallinnon on ymmärrettävä datan hallintaan liittyvät liiketoimintatarpeet, jotta sen tarjoamat tietopalvelut vastaavat liiketoiminnan tarpeita.

## ABSTRACT

**TOMMI HIRVONEN:** Improving research data management and assessing data quality in an expert organisation

Tampere University of Technology

Master of Science Thesis, 90 pages, 8 Appendix pages

June 2016

Master's Degree Programme in Information and Knowledge Management

Major: Information Management and Systems

Examiner: Professor Samuli Pekkola

**Keywords:** data management, data quality assessment, data governance

The quality of the reports and decisions made by a company is defined by the quality of the data it uses. Data quality is an important factor in a company's business, but the huge amount of data in modern businesses places challenges to data management. The good quality of data can be ensured both short-term and long-term through effective data management. However, data management is not just a technical problem. In addition to technology, both people and processes are fundamentally linked to data management. Taking these organisational aspects into account when managing data is essential.

In this thesis the methods with which a company can improve their data management and assess data quality were researched. The company studied in this thesis is a Finnish medium-sized expert company and the studied data was the company's research data. Theory regarding data quality, data quality assessment, data governance as well as data management and quality improvement was researched with a literature review. The company's research data quality was assessed using a subjective and qualitative interview method. In addition, problems related to the company's data management were surveyed using interviews. Based on the discovered problems the current level of the company's data management maturity was defined as well as a target level. Finally, improvement suggestions were formed to help the company improve its data management toward the target level.

Data quality dimensions are properties of data which can be measured. These dimensions can be assessed using established methods utilising objective and/or subjective metrics. In this thesis three methods for data quality assessment were introduced in detail. Companies can use these tools to assess the quality of their data.

Companies can improve their data management by utilising frameworks of data governance among other resources. Data governance is the definition of roles, responsibilities, practices and processes to ensure effective data management. On the other hand, data stewardship ensures that the strategic policies of data governance are transformed into a tactical plan and implemented in day-to-day working. When improving data management, it is essential to identify the critical success factors affecting the success of the project. Literature also offers data management best practices which help companies to improve their business. It is also important to ensure that data management improvement has the necessary resources. They make improvement possible, but they have to be utilised correctly. The co-operation of the business department and the IT department is essential when improving data management. The IT department must understand the company's business needs in order to offer relevant information services.

## ALKUSANAT

Vanha klisee osoittautui kohdallani sitä todemmaksi, mitä pidemmälle pääsin diplomityöni kirjoitusprosessissa; hyvin suunniteltu todellakin on puoliksi tehty. Työn huolellinen suunnittelu siivitti itse työn tekemistä, jolloin diplomityön kirjoittaminen oli sujuvaa ja mukaansatempaavaa. En osannut odottaa, että diplomityön tekeminen voisi olla näinkin kiehtovaa ja innostavaa. Nyt ymmärrän paremmin niitä, jotka haluavat luoda uraa tutkijana. Tämähän on todella kiinnostavaa työtä!

Työn tekemistä helpotti myös itselleni sopiva ja erittäin kiinnostava aihe. Aiheessa yhdistyivät loistavasti omat kiinnostukseni kohteet sekä kohdeyrityksen tarpeet. Diplomitöiden kohdalla työn tekijän ja yrityksen kiinnostukset eivät aina kohtaa, joten ymmärrän omalla työlläni olleen ideaalinen lähtötilanne. Toivottavasti työstäni on kohdeyritykselle hyötyä ja työni vaikuttaa yrityksen toimintatapaan tai ajatusmaailmaan ainakin jossain määrin.

Vaikka olen diplomityön jälkeen siirtynyt toisaalle, muistelen erityisellä lämmöllä aikaani kohdeyrityksessä ennen diplomityötä ja sen aikana. Haluankin kiittää koko yritystä lämpimästä, osallistavasta ja tukevasta kulttuurista – työkaverit tekevät työpaikan. Erityisesti haluan kiittää esimiestäni, joka aikoinaan huomasi ja näki arvon ehdottamassani tutkimusaiheesta. Suuri kiitos kuuluu myös kaikille työn haastatteluihin osallistuneille. Ilman heidän panostaan työ olisi jäänyt vaillinaiseksi. Lisäksi kiitän kaikkia niitä, jotka kahvihuonekeskusteluissa kiinnostuivat työstäni ja jotka työn loppua kohden kiinnostuivat myös sen tuloksista.

Ohjaajaani kiitän hänen arvokkaista näkemyksistään ja kehitysehdotuksistaan, jotka auttoivat tekemään työstä toimivan kokonaisuuden. Kiitän myös opiskelukavereitani työhön liittyvien ajatusten kommentoinnista sekä vanhempiani työn oikoluvusta ja jatkuvasta tuesta.

Espoossa, 16.6.2016

Tommi Hirvonen

# SISÄLLYSLUETTELO

1.	JOHDANTO .....	1
1.1	Kohdeyritys .....	1
1.2	Tutkimuksen perustelu ja merkitys .....	2
1.3	Tutkimusongelma .....	3
1.4	Tutkimusvalinnat .....	4
1.5	Rakenne .....	5
2.	DATAN LAATU JA SEN ARVIOINTI .....	7
2.1	Laadun ulottuvuudet .....	7
2.2	Laadun arvioinnin menetelmien vertailukohdat .....	9
2.3	Laadun arvioinnin menetelmät .....	12
2.3.1	AIMQ-menetelmä .....	13
2.3.2	Total Data Quality Management .....	14
2.3.3	IPMAP .....	16
2.4	Yhteenveto .....	18
3.	DATAN HALLINNOINTI .....	20
3.1	Hallinnoinnin tasot .....	21
3.2	Hallinnoinnin roolit .....	25
3.3	Hallinnoinnin viitekehykset ja prosessi .....	27
3.4	Isännöinti osana datan hallinnointia .....	33
3.5	Yhteenveto .....	34
4.	DATAN HALLINNAN JA LAADUN KEHITTÄMINEN .....	36
4.1	Viitekehys datan hallinnan ongelmien löytämiseen .....	36
4.2	Hallinnoinnin ja laadun kehittämisen kriittiset onnistumistekijät .....	37
4.3	Organisatoriset käytännöt datan laadun parantamiseksi .....	39
4.4	Datan laadun kehittämisen prosessi .....	42
4.5	Yhteenveto .....	45
5.	MENETELMÄT .....	47
5.1	Datan laadun arviointi .....	47
5.2	Datan hallinnan arviointi .....	50
6.	TULOKSET .....	52
6.1	Datan laadun arvioinnin tulokset .....	53
6.2	Laadun arvioinnin tulosten tiivistys .....	61
6.3	Datan hallinnan arvioinnin tulokset .....	64
7.	TULOSTEN POHDINTA JA KEHITYSEHDOTUKSET .....	68
7.1	POSMAD-viitekehys .....	68
7.2	Muut huomiot tuloksista .....	70
7.3	Datan hallinnan nykytaso .....	71
7.4	Datan hallinnan tavoitetaso .....	72
7.5	Kehitysehdotukset .....	75
8.	PÄÄTELMÄT .....	81

LÄHTEET.....	86
--------------	----

LIITE A: HAASTATTELUPOHJA

LIITE B: AIMQ-KYSELYLOMAKE

LIITE C: POSMAD-VIITEKEHYS

LIITE D: DATAN HALLINNAN ONGELMAT

# 1. JOHDANTO

Datan määrä on nykyaikaisissa yrityksissä valtava, ja useimmissa organisaatioissa datan määrän arvioidaan kasvavan 35 – 50 % joka vuosi (Beath et al. 2012, s. 1). Yrityksissä luodun datan määrän suuri kasvu asettaa haasteita datan laadun ylläpitämiselle, mutta samalla tekee siitä entistä tärkeämpää, sillä tätä dataa hyödynnetään liiketoimintapäätöksissä (Storey et al. 2012, s. 1). Datan laatu vaikuttaa muun muassa yrityksen rahoituspäätöksiin, dokumentaatiokäytäntöihin sekä organisaation suhteisiin (Sebastian-Coleman 2013, s. xxxii). Datan laadusta on siis tullut yhä merkittävämpi tekijä yrityksen liiketoiminnalle (Lee et al. 2002, s. 1; Wang et al. 2008, s. 1; Laatikainen & Niemi 2012, s. 1). Huonolaatuinen data voikin pahimmillaan aiheuttaa ongelmia ja menetyksiä yritykselle (Laatikainen & Niemi 2012, s. 1; Storey et al. 2012, s. 1). Toisaalta datan laatu myös määrittelee, kuinka hyviä yrityksen tekemät raportit ja päätökset voivat olla (Cheong & Chang 2007, s. 1000). Tämän vuoksi erityisesti tietointensiivisten yritysten tulee keskittyä datan hallintaan ja varmistaa, että yrityksen data täyttää sille asetetut vaatimukset. Näin dataa voidaan todella käyttää tukemaan yrityksen tekemiä analyysejä sekä liiketoiminnallista päätöksentekoa.

Yritysten tulee myös ymmärtää, että datan laadun hallinta ei ole ainoastaan tekninen ongelma. Usein yritykset antavat vastuun datan laadun hallinnasta IT-osastoille, jolloin laadun hallinnalle kriittiset organisatoriset ongelmat jätetään huomioimatta (Weber et al. 2009, s. 1). Datan hallinnan hankkeissa usein neljä viidesosaa tekemisestä liittyy ihmisiin ja prosesseihin ja loput varsinaiseen tekniseen toteutukseen (Niemi & Kontra 2012, s. 3-4). Kehittääkseen datan laatua, vaatii tietojärjestelmien inhimillinen näkökulma yhtä paljon huomiota kuin niiden tekninen puoli (Storey et al. 2012, s. 440). Jotta yritys voi hallita dataansa ja sen laatua tehokkaasti, on olennaista huomioida teknisten aspektien lisäksi myös organisatoriset näkökulmat datan hallintaan.

## 1.1 Kohdeyritys

Tässä diplomityössä tutkitaan, miten asiantuntijayrityksen tutkimusdatan hallintaa ja datan laatua voidaan kehittää. Työn kohdeyritys on suomalainen keskikokoinen asiantuntijayritys, joka tekee muun muassa kemian, hydrologian ja geologian tutkimustyötä. Kohdeyrityksessä on noin sata työntekijää, joista suuri osa työskentelee asiantuntijatehtävissä. Yrityksen tutkimusdataa käyttää lähes koko organisaatio eri tarkoituksiin. Tutkimusdata sijaitsee sen pitkäaikaiseen säilytykseen suunnitellussa tietokannassa, ja suuri osa datasta

on kemian laboratoriotuloksia sekä kemiaan ja hydrologiaan liittyvää sensoridataa. Tietokanta sisältää myös paljon muuta dataa, mutta näihin ei keskitytä tämän työn yhteydessä.

Tutkimusdatan erityispiirre kohdeyrityksen kontekstissa on datan vahva numeerisuus. Data koostuu siis suurilta osin luvuista ja lyhyistä merkkijonoista. Tutkimusdatalla ja esimerkiksi yleisesti asiakasdatalla on kuitenkin yhteistä se, että muun muassa tietojen eheydellä on todella suuri merkitys. Siinä missä asiakasdatan osoitetiedot tulee ilmaista aina samalla tavalla, on tutkimusdatassa kohdetiedot ilmoitettava aina samoin. Tutkimusdatan laadulla on kohdeyrityksessä suuri merkitys, sillä dataa käytetään hyödyksi muun muassa yrityksen tekemissä tutkimusraporteissa. Dataa käytetään analyysihin, jotka vaikuttavat yrityksessä tehtäviin päätöksiin.

## **1.2 Tutkimuksen perustelu ja merkitys**

Työn aihe on ajankohtainen, sillä datan nopeasti kasvava määrä yrityksissä vaatii datan hallintaan keskittymistä. Erityisesti kohdeyrityksen tapauksessa datan hallinta on ajankohtainen toiminto, sillä jo yksistään datan valtava määrä asettaa haasteita sen hallinnalle. Työn tärkeimmät tulokset kohdeyrityksen kannalta ovat datan hallinnan kehitysehdotukset.

Työn tarkoituksena ei ole tuottaa uusia näkemyksiä tai luoda uusia synteesejä datan hallinnan alalla. Sen sijaan työ keskittyy datan hallinnan teoriaan olemassa olevan kirjallisuuden avulla ja kohdeyrityksen nykytilanteen empiiriseen tutkimiseen. Työssä esitelty teoria auttaa ja tukee tutkimuksen empiiristä osaa. Tieteellinen kontribuutio datan hallinnan alalle voikin jäädä pieneksi näistä syistä johtuen. Työssä kuitenkin sovelletaan datan hallinnan teoriaa sekä alan vallitsevia parhaita käytäntöjä ja datan laadun arvioinnin menetelmiä. Työn lopussa arvioidaan, kuinka hyvin nämä käytännöt ja menetelmät toimivat käytännössä ja kuinka helppo niitä on soveltaa tämän työn kontekstissa kohdeyrityksen toimintaan.

Työn käytännöllinen kontribuutio on puolestaan suurempi. Parhaimmassa tapauksessa työstä on kohdeyritykselle merkittävää liiketoiminnallista hyötyä. Sen avulla yritys voi parantaa omia datan hallinnan käytäntöjä sekä datan laatua. Tutkimuksen onnistumista arvioidaan pohtimalla kuinka hyvin saadut tulokset vastaavat tutkimuskysymyksiin. Työn henkilökohtainen merkitys tekijälle on myös suuri, sillä se toimii erinomaisena itseopiskelun keinona ja mahdollisuutena soveltaa oppeja ja teoriaa liiketoiminnan käytännön ongelmiin. Työstä voi olla myös hyötyä muille yrityksille, jotka haluavat kehittää omaa datan hallintaa tai kartoittaa oman yrityksensä datan laadun nykytilannetta.



### 1.3 Tutkimusongelma

Tässä tutkimuksessa selvitetään, miten asiantuntijayrityksen tutkimusdatan hallintaa voidaan kehittää. Työssä arvioidaan kohdeyrityksen datan hallinnan ja datan laadun nykytilaa sekä luodaan kehitysehdotuksia näiden kehittämiseen. Työ keskittyy haasteiden ja epäkohtien löytämiseen kohdeyrityksen datan hallinnassa. Tarkasteltava data tullaan rajaamaan asiantuntijoiden käyttämään tutkimusdataan ja vielä tarkemmin kemian sekä hydrologian laboratoriotuloksiin ja sensoridataan. Tämä data sijaitsee lähinnä tutkimusdatatietokannassa. Tarkastelua ei kuitenkaan rajata vain tietokantaan, sillä osa käytetystä tutkimusdatasta sijaitsee muualla, esimerkiksi taulukkolaskentaohjelmilla luoduissa taulukoissa työntekijöiden tietokoneilla.

Lisäksi tutkimuksessa arvioidaan asiantuntijayrityksen tutkimusdatan laatua. Työssä karroitetaan kohdeyrityksen kemian ja hydrologian tutkimusdatan laadun nykytilaa. Tässäkään tapauksessa tarkastelua ei rajata yksinomaan tietokantaan edellisessä kappaleessa mainituista syistä johtuen. Datan laadun arvioinnissa tullaan hyödyntämään kirjallisuudessa (Lee et al. 2002) esiteltyä AIMQ-menetelmää soveltuvien osien.

Työn teoriaosassa käsitellään datan laatua, sen arviointia sekä datan hallinnan teoriaa. Teoriaosan tarkoitus on tukea tutkimuksen empiiristä osaa ja kehitysehdotusten tekemistä. Teoriaosassa esitellään datan laadun näkökulmia ja datan laadun arviointimenetelmiä sekä käsitellään datan hallinnan käytäntöjä ja merkitystä yritykselle. Lisäksi teoriassa esitellään työssä käytetty menetelmä datan laadun arviointiin.

Tämän diplomityön päätutkimuskysymys on:

- *Miten yrityksen datan hallintaa voidaan kehittää ja datan laatua arvioida?*

Työn alakysymykset ovat:

- *Mikä on kohdeyrityksen tutkimusdatan laadun nykytilanne?*
- *Mitä haasteita ja epäkohtia liittyy kohdeyrityksen tutkimusdatan hallintaan?*
- *Mikä on kohdeyrityksen tutkimusdatan hallinnan nykytaso ja mikä on sen tavoitetilaa?*
- *Miten kohdeyrityksen tutkimusdatan hallintaa voidaan kehittää kohti tavoitetilaa?*

Työn päätutkimuskysymykseen vastataan yleisellä tasolla. Näin tämän työn tuloksia voidaan mahdollisesti soveltaa muissakin organisaatioissa. Päätutkimuskysymykseen vastataan teorialukujen 2-4 avulla, mutta siihen palataan myös päätelmissä.

Tutkimuksen alakysymykset liittyvät spesifisesti kohdeyrityksen toimintaan. Ensimmäiseen ja toiseen alakysymykseen pyritään vastaamaan empirialuvun 6 avulla. Ensimmäisen alakysymyksen avulla voidaan selvittää kohdeyrityksen datan laadun lähtötilanne. Tähän kysymykseen liittyvät tulokset voivat toimia referensseinä tulevaisuudessa tehdyille laadun arvioinneille. Toinen kysymys pyrkii kartoittamaan tämänhetkiset ongelmat

datan hallinnassa. Tähän kysymykseen vastaaminen auttaa myös viimeiseen kysymykseen vastaamisessa. Kolmanteen ja neljanteen alakysymykseen pyritään vastaamaan luvussa 7, jossa yhdistetään sekä teoriassa että empiriassa saatuja tuloksia. Kolmannen alakysymyksen avulla voidaan selvittää, mikä on datan hallinnan nykytaso sekä millaista tilannetta tulevaisuudessa tavoitellaan. Tämä kysymys auttaa myös sopivien kehitystoimenpiteiden valinnassa. Viimeinen alakysymys on kohdeyrityksen kannalta tärkein kysymys, sillä se vastaa juuri kohdeyrityksen tarpeeseen.

## 1.4 Tutkimusvalinnat

Taulukossa 1 on esitetty tutkimuksen luonteen kannalta tärkeimmät valinnat ja näkökulmat. Näiden valintojen pohtiminen ja huomioiminen on tärkeää, sillä ne vaikuttavat tutkijan tekemiin oletuksiin sekä tutkimuksen luonteeseen.

*Taulukko 1. Tutkimuksessa tehdyt valinnat ja näkökulmat*

Näkökulma	Tehty valinta	Vaikutus tutkimukselle
Filosofia	Tulkinnallinen	Tutkimukseen osallistuvien subjektiivisuus ja tutkittavan ilmiön organisatoriset näkökulmat huomioidaan
Lähestymistapa	Teorialähtöinen / deduktiivinen	Olemassa olevaa teoriaa käytetään analyysien tekemiseen
Tutkimusote/strategia	Tapaustutkimus	Tutkitaan yksittäistä tapausta / rajattua kokonaisuutta, kiinnostuksen kohteena prosessit ja käytännöt
Ajallinen valinta	Läpileikkaava	Tietty ilmiö tiettyyn aikaan
Tiedonkeruu	Laadullista	Haastattelut
Tiedonanalysointi	Laadullista	Sisällönanalyysi
Johtopäätösten teko	Teorian ja tulosten yhdistämistä	Tulokset ovat kehitysehdotusten muodossa

Työn tutkimusfilosofia on tulkinnallinen. Tulkinnallisuudessa korostetaan ihmisten roolia toimijoina sekä osana tutkimusta. Tutkijan tehtävänä on ymmärtää ilmiö tutkittavien henkilöiden näkökulmasta. Tulkinnallisessa tutkimuksessa aineiston keräys on laadullista ja otokset ovat pieniä. (Saunders et al. 2009, s. 116.) Työn lähestymistapa taas on teorialähtöinen eli deduktiivinen. Deduktiivisessa tutkimuksessa tutkimusaineiston analyysi perustuu olemassa olevaan teoriaan (Tuomi & Sarajärvi 2002; Saaranen-Kauppisen & Puusniekan 2006 mukaan). Tässä tutkimuksessa olemassa olevaa datan hallinnan ja datan laadun arvioinnin teoriaa käytetään kohdeyrityksen nykytilanteen selvittämiseen ja kehitysehdotusten tekemiseen. Tämän vuoksi tulkinnallisuus ja deduktiivinen lähestymistapa ovat sopivia valintoja tälle työlle.

Työn tutkimusotteeksi on valittu tapaustutkimus. Tapaustutkimuksessa tutkitaan yksittäistä tapausta, tilannetta tai rajattua kokonaisuutta, jossa kiinnostuksen kohteena ovat prosessit ja käytännöt (Saaranen-Kauppinen & Puusniekka 2006). Tämä ote sopii hyvin

tähän tutkimukseen, sillä tavoitteena on selvittää kohdeyrityksen tämänhetkiset datan hallinnan käytännöt ja kehittää yrityksen datan hallintaa. Työssä käytetään yhtä laadullista menetelmää aineiston keräämiseen, haastattelua. Tutkimuksen ajallinen valinta on puolestaan läpileikkaava. Läpileikkaavassa tutkimuksessa tarkastellaan tiettyä ilmiötä tietyllä ajanhetkellä (Saunders et al. 2009, s. 155). Tässä työssä tarkastellaan juuri kohdeyrityksen datan hallintaa ja datan laatua vuosien 2015 ja 2016 vaihteessa.

Työssä saatuja tuloksia ja aineistoa käsitellään sisällönanalyysin keinoin. Sisällönanalyysissä aineistoa tutkitaan eritellen, tiivistäen sekä yhtäläisyyksiä ja eroja etsien. Siinä tarkastellaan tekstimuotoisia tai tekstiksi muutettua aineistoa. (Tuomi & Sarajärvi 2002; Saaranen-Kauppinen & Puusniekan 2006 mukaan.) Tässä työssä tutkittavana aineistona ovat haastatteluiden tulokset sekä teoriataustassa hyödynnetty kirjallisuus.

Työn teoriaosa tehdään kirjallisuuskatsauksena, ja työhön valittu empiirinen menetelmä on haastattelu. Datan laadun arviointi suoritetaan soveltuvin osin AIMQ-menetelmän (Lee et al. 2002) mukaan. AIMQ-menetelmä on käsitelty luvussa 2.3.1. Laadun arvioinnissa tullaan tarkastelemaan mahdollisimman monta datan laadun ulottuvuutta. Ulottuvuudet valitaan kuitenkin niin, että niiden keskinäiset erot ovat haastateltaville selkeät. Haastattelu soveltuu tähän tarkoitukseen paremmin kuin esimerkiksi kysely, sillä haastattelussa haastattelija voi täsmentää eri ulottuvuuksien määritelmiä. Datan hallinnan haasteiden ja epäkohtien selvittämisessä käytetään myös haastattelua. Datan hallintaan liittyvät kysymykset tullaan esittämään datan laadun arvioinnin kanssa samassa haastattelussa. Tulosten analysointiin käytetään POSMAD-viitekehystä (McGilvray 2008). Työssä käytetyt empiiriset menetelmät on esitelty tarkemmin luvussa 5.

Puolistrukturoidut haastattelut mahdollistavat haastattelijan ja haastateltavan välisen vuorovaikutuksen, jolloin haastattelija voi esittää täsmentäviä kysymyksiä, ja asioihin voidaan perehtyä syvällisemmin haastatteluiden aikana. Haastattelut tullaan pitämään yksilöhaastatteluina. Haastateltavat valitaan harkintaan pohjautuvalla otannalla. Harkintaan pohjautuvassa otannassa valitaan henkilöt, jotka pystyvät parhaiten vastaamaan tutkimuskysymyksiin (Saunders et al. 2009, s. 237). Haastatteluun valitaan kohdeyrityksen työntekijöitä, jotka omassa toimenkuvassaan työskentelevät tutkimusdatan kanssa sen elinkaaren eri vaiheissa. Tutkimusdataa työkseen tuottavat, ylläpitävät ja käyttävät työntekijät osaavat parhaiten vastata dataa koskeviin kysymyksiin. Tämän vuoksi oikeiden henkilöiden valitseminen tutkimukseen on tärkeä vaihe.

## 1.5 Rakenne

Työn rakenne on seuraavanlainen. Luvut 2-4 ovat teorialukuja, joissa esitellään työn kannalta tärkeä teoriatausta. Luvussa 2 käsitellään datan laadun käsitettä sekä datan laadun arvioinnin menetelmiä. Luvussa 3 esitellään datan hallinnointia, mitä se on ja mitä se sisältää. Luku 4 keskittyy datan hallinnan ja laadun kehittämisen kriittisten osa-alueiden esittelyyn.

Luvussa 5 esitellään työssä käytetyt empiiriset menetelmät sekä empiirisen osan vaiheet. Työssä on kaksi empiiristä kokonaisuutta: datan laadun arviointi ja datan hallinnan arviointi. Luvussa 6 käsitellään sekä datan laadun että datan hallinnan arvioinnissa saadut tulokset. Luvussa 7 tehdään päätelmiä saatujen tulosten sekä työssä esitellyn teorian pohjalta. Lisäksi luvussa määritellään kohdeyrityksen datan hallinnan kypsyyden nyky- ja tavoitetaso sekä luodaan kehitysehdotukset datan hallinnan kehittämiseksi.

Liitteessä A on esitelty työssä käytetty haastattelupohja, joka perustuu liitteessä B esitettyyn AIMQ-kyselylomakkeeseen. Liitteessä C on esitetty POSMAD-viitekehys, joka on käsitelty myös luvussa 4.1. Liitteessä D on POSMAD-viitekehysten matriisiin koottu kohdeyrityksessä havaitut datan hallinnan ongelmat.

## 2. DATAN LAATU JA SEN ARVIOINTI

Datan laatua käsittelevässä kirjallisuudessa käytetään usein sanoja data, tieto ja informaatio keskenään vaihtokelpoisesti (Wang et al. 2008, s. 1). Samoin tässä työssä käytetään edellä mainittuja sanoja vaihtokelpoisesti ottamatta kantaa sanojen merkitysten lieviin eroihin tai tiedon eri tasoihin. Työssä käytettyä aineistoa käännettäessä englannin kielen sana ”data” kääntyy luonnollisesti dataksi, mutta ”information” on käännetty joko tiedoksi tai informaatioksi.

Datan laadusta on vaikea antaa yleispätevää määritelmää, sillä käsitys laadusta vaihtelee kontekstista toiseen. On tehtävä oletuksia siitä, mitkä laadun näkökulmat pätevät missäkin tilanteessa. Nimenomaisesti datan laadussa voidaan tarkastella laadun eri piirteitä, jotka ovat juuri dataa tarkasteltaessa tärkeitä ja relevantteja. (Bobrowski et al. 1999, s. 5.) Datan laadulla voidaan muun muassa tarkoittaa sitä, saavuttaako data sen käyttäjien epäsuorat tai selkeät odotukset tai vaatimukset. Laatu siis riippuu siitä, mitä sen käyttäjät siltä odottavat. (Sebastian-Coleman 2013, s. 40.) Toisen määritelmän mukaan datan laadulla viitataan siihen asteeseen, jolla dataa voidaan pitää luotettavana lähteenä vaadittuun käyttöön (McGilvray 2008, s. 5). Datan laadun mittaaminen taas riippuu siitä, kenen ja minkä roolin näkökulmasta dataa tarkastellaan, sillä eri organisaatio-rooleilla voi olla eriäviä näkemyksiä samasta datasta ja sen laadusta (Caballero et al. 2007, s. 1-2).

Datan laadun määrittelyä voidaan lähestyä niin sanottujen laadun ulottuvuuksien kautta. Niiden avulla voidaan määrittellä, mitä datan laadulla tarkoitetaan missäkin kontekstissa. Ulottuvuudet myös mahdollistava datan laadun mittaamisen, kun voidaan määrittellä, mitä asioita datan laadusta mitataan. Datan laatu voi koostua useista laadun ulottuvuuksista.

Edellä mainituista syistä johtuen datan laatu voidaan nähdä subjektiivisena asiana. Datan koettu laatu voi vaihdella huomattavasti esimerkiksi eri käyttäjien ja roolien välillä. Tämä osaltaan hankaloittaa datan laadun objektiivista arviointia, mutta kuten tässä luvussa huomataan, ei objektiivisten mittareiden käyttäminen ole itseisarvo datan laadun arvioinnissa.

### 2.1 Laadun ulottuvuudet

Jotta datan laatua voidaan mitata, tulee laadulle ensin määrittellä mittarit. Mittarit ovat jonkin ominaisuuden tai piirteen kvantitatiivinen esitystapa. Toisin sanoen, mittarit kvantifioivat niitä asioita, joita datan laadussa mitataan. Datan laadun ulottuvuudet ovat datan osa-alueita, joita voidaan mitata. (Caballero et al. 2007, s. 3; Sebastian-Coleman 2013, s. 47.) Toisen määritelmän mukaan datan laadun ulottuvuus on kokoelma datan laadun piirteitä, jotka esittävät yhden näkökulman datan laadusta (Wang & Strong 1996, s. 6). Datan laadun ulottuvuudet ovat tärkeitä, koska ne auttavat ymmärtämään miksi dataa mitataan

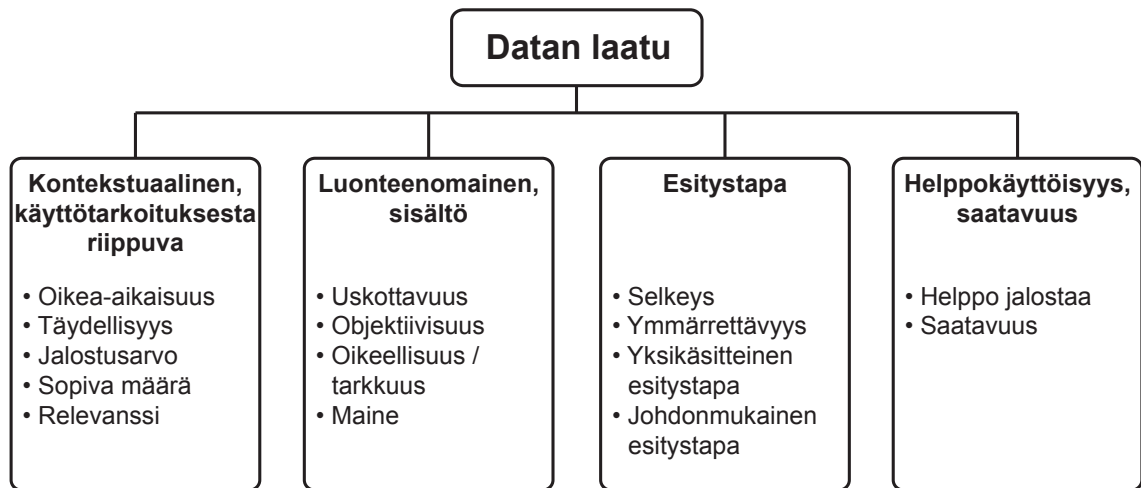
ja arvioidaan (Sebastian-Coleman 2013, s. 47). Taulukossa 2 on esitetty datan laadun ulottuvuuksia sekä missä teoksissa kyseiset ulottuvuudet on esitetty.

Datan laadun mittarit puolestaan määrittelevät tietyn datan, jota mitataan sekä mitä siitä mitataan (Sebastian-Coleman 2013, s. 47). Yhteen ulottuvuuteen voi liittyä useampi mittari (Batini et al. 2009, s. 6). Datan laadun ulottuvuudet vastaavat toisin sanoen kysymykseen miksi, laadun arvioinnin menetelmät kysymykseen miten ja laadun mittarit kysymykseen mitä (Sebastian-Coleman 2013, s. 48).

**Taulukko 2.** Datan laadun ulottuvuudet

Ulottuvuus	Kuvaus	Sebastian-Coleman 2013, s. 62	Bobrowski et al. 1999, s. 5	Lee et al. 2002, s. 137
Täydellisyys	Data ei ole puutteellista	X	X	X
Oikea-aikaisuus	Data on ajankohtaista, ei vanhentunutta	X	X	X
Oikeellisuus	Kuinka hyvin data vastaa standardeja	X		
Eheys	Datan sisäinen johdonmukaisuus, sama tieto on aina esitetty samalla tavalla, datan osat eivät ole ristiriidassa keskenään	X	X	
Käyttäjien odotuksiin vastaaminen	Korkealaatuinen data vastaa paremmin odotuksia kuin huonolaatuinen data	X		
Relevanssi	Jokaisen datan osan säilöminen on tärkeää todellisuuden kuvaamiseksi		X	X
Luotettavuus	Dataan voidaan luottaa		X	
Tarkkuus	Data kuvaa todellisuuden ilmiötä riittävän tarkasti		X	
Yksikäsitteisyys	Jokaisella datan osalla on uniikki merkitys		X	
Täsmällisyys	Data on esitetty riittävällä tarkkuudella		X	
Objektiivisuus	Data ei riipu ihmisten tekemistä tulkinnoista		X	X
Tiiviys	Tarvittavan datan ohella ei ole ylimääräistä dataa		X	X
Käytettävyys	Dataa voidaan käyttää suunnitellusti		X	X
Hyödyllisyys	Jos data on hyvälaatuista, sitä käytetään ja hyödynnetään päätöksenteossa		X	X
Sopiva määrä	Dataa ei ole liikaa tai liian vähän organisaation tarpeisiin nähden			X
Turvallisuus	Pääsynvalvonta, autentikointi			X
Ymmärrettävyys	Dataa on helppo ymmärtää			X
Saatavuus	Data on helposti saatavissa			X
Helppokäyttöisyys	Dataa on helppo jalostaa omiin tarpeisiin sopivaksi			X
Maine	Datan tunnetaan olevan laadukasta			X

Datan laadun ulottuvuudet voidaan luokitella neljään kategoriaan. Ulottuvuus voi liittyä datan luonteeseen, helppokäyttöisyyteen ja saatavuuteen, esitystapaan tai datan kontekstin tai käyttötarkoitukseen. (Wang & Strong 1996; DGIQ 2007; Rickards & Ritsert 2012.) Kuvassa 1 on esitetty yleisimpiä datan laadun ulottuvuuksia sekä niiden luokittelu.



**Kuva 1.** Datan laadun ulottuvuudet ja niiden luokittelu (mukaillen Wang & Strong 1996; DGIQ 2007; Rickards & Ritsert 2012)

*Luonteenomaiset* ulottuvuudet kuvaavat sitä, että datalla on omasta takaa tiettyjä ominaisuuksia. *Kontekstuaaliset* ulottuvuudet puolestaan ilmaisevat sitä, että datan laatua tulee tarkastella tietyssä kontekstissa (aika, paikka, käyttäjät jne.). *Esitystapaan* ja datan *saatavuuteen* liittyvät ulottuvuudet korostavat järjestelmien merkitystä. Järjestelmien tulee olla helposti saatavilla, mutta samalla turvallisia. Lisäksi niiden tulee esittää data muodossa, joka on helppo ymmärtää ja tulkita. (Wang & Strong 1996, s. 19.) Tällaiset luokitellut, useiden ulottuvuuksien luettelot auttavat laadun arvioinnissa sekä todentamaan laadun kehitystoimenpiteiden tehokkuutta. Vaarana kuitenkin on, että tällöin datan laadun määritelmä muuttuu liian laajaksi ja kaikkia keinoja datan käsittelyyn pidetään laadunvarmistusmenetelminä (esimerkiksi uuden työkalun käyttöönotto). (Rickards & Ritsert 2012, s. 29.) On siis varmistettava, että spesifisesti datan laadun varmistamiseksi ja parantamiseksi suunnitellut menetelmät ja työkalut toimivat laadunvarmistustoimenpiteinä. Datan prosessointi ei välttämättä tarkoita, että datan laatu paranee käsittelyn myötä.

## 2.2 Laadun arvioinnin menetelmien vertailukohdat

Datan laadun arviointiin on olemassa useita vakiintuneita menetelmiä (Batini et al. 2009, s. 1). Nämä menetelmät tarjoavat viitekehyksiä, joiden avulla datan laadun arviointi voidaan suorittaa systemaattisesti ja tehokkaasti. Datan laadun arvioinnin menetelmiä voidaan vertailla useiden eri näkökulmien kautta. Taulukkoon 3 on koottu eri näkökulmat ja niihin liittyvät yksityiskohdat. Tässä luvussa tarkastellaan näitä vertailukohtia tarkemmin.

**Taulukko 3.** *Datan laadun arvioinnin menetelmien vertailunäkökulmat (Batini et al. 2009, s. 3-11)*

Vaiheet	Strategiat	Ulottuvuudet ja mittarit	Laatuun liittyvät kustannukset	Tarkastellut datatyytit	Tarkastellut tietojärjestelmätyypit
<ul style="list-style-type: none"> <li>• data-analyysi</li> <li>• laatuvaatimusanalyysi</li> <li>• kriittisten alueiden tunnistaminen</li> <li>• prosessimallinnus</li> <li>• laadun mittaaminen</li> </ul>	<ul style="list-style-type: none"> <li>• datalähtöinen</li> <li>• prosessilähtöinen</li> </ul>	<ul style="list-style-type: none"> <li>• objektiiviset mittarit</li> <li>• subjektiiviset mittarit</li> </ul>	<ul style="list-style-type: none"> <li>• datan laadun ohjelman kustannukset</li> <li>• huonoon datan laatuun liittyvät kustannukset</li> </ul>	<ul style="list-style-type: none"> <li>• rakenteellinen data</li> <li>• rakenteeton data</li> <li>• semirakenteellinen data</li> </ul>	<ul style="list-style-type: none"> <li>• monoliittiset</li> <li>• datavarastot</li> <li>• hajautetut</li> <li>• yhteistoiminnalliset</li> <li>• verkkotietojärjestelmät</li> <li>• P2P-tietojärjestelmät</li> </ul>

Batini et al. (2009, s. 4) mukaan datan laadun arviointiprosessin vaiheet ovat:

- *data-analyysi*, jossa pyritään ymmärtämään dataa ja sen hallinnan käytäntöjä
- *laatuvaatimusanalyysi*, jossa haastatellaan datan käyttäjiä ja ylläpitäjiä laatuongelmien tunnistamiseksi ja uusien tavoitteiden asettamiseksi
- *kriittisten alueiden tunnistaminen*, jossa valitaan olennaisimmat datapaketit, jotka sisältyvät arviointiin
- *prosessimallinnus*, jonka avulla luodaan malli dataan tuottamiseen liittyvästä prosessista
- *laadun mittaaminen*, jossa valitaan laadun ulottuvuudet, joihin tunnistetut ongelmat vaikuttavat.

Jotkin menetelmät sisältävät arviointiprosessin lisäksi viitekehyksen datan kehittämiseksi. Kehitysprosessin vaiheet Batini et al. (2009) mukaan on kuvattu myöhemmin luvussa 4.3. Lisäksi Bobrowski et al. (1999, s. 9) esittävät kymmenen vaihetta datan laadun arvioinnille:

1. valitse kiinnostavat laadun ulottuvuudet, kaikki ulottuvuudet eivät ole relevantteja
2. valitse tai määrittele kysymykset, jotka kuvailevat valittuja ulottuvuuksia
3. valitse tai määrittele mittarit ja tekniikat vastataksesi kaikkiin kysymyksiin
4. määrittele jokaiselle mittarille arvot tai arvoalueet, jotka kuvaavat hyvälaatuista tai huonolaatuista dataa
5. jos olet valinnut subjektiiviset mittarit, määrittele sopivat kyselylomakkeet ja tiedonkeruumenetelmät
6. sovelta valittuja tekniikoita vastaaviin kohteisiin (datamalli, tietokanta, dokumentaatio jne.)
7. kerää tieto kyselylomakkeilla
8. arvioi jokaisen mittarin kohdalla onko datan laatu hyväksyttävää vai ei, tee korjaustoimenpiteet
9. talleta saadut tulokset
10. palaa vaiheeseen 1.



Edellä esitetty kymmenvaiheinen prosessi laadun arvioinnille vastaa Batini et al. (2009) prosessin viimeistä vaihetta: laadun mittaaminen. Bobrowski et al. (1999) malli ei ota huomioon laadun arviointiin liittyvä taustatyötä, esimerkiksi data-analyysiä ja prosessimallinnusta. Sen viimeinen vaihe kuitenkin korostaa laadun arvioinnin iteratiivisuutta ja jatkuvaa datan hallintaa. Prosessia siis jatketaan niin kauan, kunnes datan laatu on määritellyissä tai hyväksyttävissä rajoissa. Esitellyt kaksi prosessimallia täydentävät toisiaan ja näin antavat kattavamman kuva laadun arvioinnin prosessista. Eri menetelmiä voidaan vertailla sen mukaan, miten ne toteuttavat kunkin arvioinnin vaiheen.

Datan laadun arvioinnin menetelmät voivat soveltaa kahta strategiaa: datalähtöinen ja prosessilähtöinen. Datalähtöiset strategiat pyrkivät kehittämään datan laatua suoraan muuttamalla dataa, kun taas prosessilähtöiset strategiat pyrkivät kehittämään datan laatua muuttamalla prosesseja, jotka tuottavat tai päivittävät dataa. (Batini et al. 2009, s. 5.) Datalähtöiset menetelmät vastaavat akuuttiin tarpeeseen korjaamalla datan, mutta jos organisaatiossa syntyy virheellistä dataa, tulisi dataan liittyviä prosesseja todennäköisesti arvioida uudelleen. Datalähtöisiä menetelmiä voidaan käyttää siinä tilanteessa, kun virheiden lähde ei voida korjata. (Loshin 2011, s. 26.) Datalähtöisissä strategioissa lähestyminen laadun kehittämiseen on lähinnä reaktiivista. Datan laadulle asetetaan kontrollit, joita seurataan. Prosessilähtöisissä strategioissa muokataan datan elinkaareen liittyviä prosesseja, jolloin on parempi mahdollisuus ottaa kantaa juurisyihin.

Arviointimenetelmät voivat tarkastella useita laadun ulottuvuuksia ja käyttää useita eri mittareita. Datan laadun arviointiin käytettyjen ulottuvuuksien ja mittarien määrittelemisen on arviointiprosessin kriittinen toiminto (Batini et al. 2009, s. 6). Tämän vuoksi käytetyt ulottuvuudet ja mittarit on valittava harkiten ja siten, että ne varmasti sopivat kohdeorganisaation kontekstiin. Lisäksi ulottuvuuksien merkitys ja niiden eroavaisuuksien tulee olla selkeitä kaikille arviointiprosessiin kuuluville henkilöille.

Laadun mittarit voidaan jakaa subjektiivisiin ja objektiivisiin mittareihin. Subjektiiviset mittarit heijastavat dataan liittyvien sidosryhmien tarpeita ja kokemuksia. Objektiiviset mittarit voivat olla joko tehtäväriippumattomia tai tehtäväriippuvia. Tehtäväriippumattomat mittarit kuvaavat datan tilaa ilman kontekstuaalista tietoa, kun taas tehtäväriippuvat mittarit kehitetään jossain sovelluskohtaisessa kontekstissa käyttäen organisaation liiketoimintasääntöjä. (Pipino et al. 2002, s. 211.) Toisen käsityksen mukaan objektiiviset mittarit perustuvat kvantitatiivisiin mittareihin, kun taas subjektiiviset mittarit perustuvat kvalitatiiviseen arviointiin (Batini et al. 2009, s. 4). Jotta datan kuluttajien näkemykset voidaan huomioida, tulee arviointimenetelmän sisältää myös subjektiivisia mittareita (Lee et al. 2002, s. 135). Objektiivinen mittaaminen sisältää vähintään toisen seuraavista vertauksista: dataa voidaan joko verrata johonkin selkeästi määriteltyyn standardiin tai itseensä ajan suhteen (Sebastian-Coleman 2013, s. 60).

Menetelmissä voidaan arvioida myös kustannuksia eri tavalla. Datan laadun kustannus on datan laadun ohjelman kustannusten ja huonolaatuiseen dataan liittyvien kustannusten

summa. Huonolaatuisen datan aiheuttamia kustannuksia voidaan periaatteessa vähentää tehokkaammilla ja kalliimmilla datan laadun ohjelmilla, jolloin taas datan laadun ohjelman kustannukset kasvavat. (Batini et al. 2009, s. 9.) On kuitenkin syytä huomata, että kyse ei ole kahden kustannuksen välillä tasapainoilusta. Jos tehokkaammalla datan laadun ohjelmalla voidaan pienentää huonolaatuisen datan aiheuttamia kustannuksia kokonaiskustannusten pysyessä samana, on tämä aina kannattavaa. Parempilaatuinen data hyödyntää koko liiketoimintaa pitkällä aikavälillä ja datan koettu arvo nousee. Huonolaatuisen datan aiheuttamia kustannuksia on kuitenkin vaikea arvioida, sillä ne ovat todella riippuvaisia datan käyttökontekstista (Batini et al. 2009, s. 9).

Datan laadun arvioinnin menetelmät voivat tarkastella yhtä tai useampaa kolmesta datatyyppistä. Rakenteellinen tieto on esimerkiksi relaatiotauluja tai statistista dataa (Batini et al. 2009, s. 9). Tietokantojen sisältämä tieto, esimerkiksi lukuarvot ja merkkijonot, on siis useimmiten rakenteellista tietoa. Rakenteeton tieto taas on dataa, joka ei ole tietokannassa, esimerkiksi sähköpostit, diaesitykset ja dokumentit (Rouse 2015). Rakenteetonta tietoa säilötään organisaatioissa usein dokumenttienhallintajärjestelmissä. Semirakenteellinen data on rakenteellista dataa, jolla on jonkinasteista joustavuutta, esimerkiksi XML-tiedostot (Batini et al. 2009, s. 9-10).

Lisäksi menetelmät voivat keskittyä erityyppisten tietojärjestelmien tarkasteluun. Monoliittiset tietojärjestelmät ovat usein tietokantoja, jotka eivät tarjoa palveluita dataan pääsemiseksi. Datavarastot ovat keskitetty kokoelma dataa, joka on kerätty useista eri lähteistä ja järjestelmistä. Hajautetut tietojärjestelmät voivat koostua useista eri tietokannoista, mutta toimivat loogisesti yhtenä järjestelmänä loppukäyttäjälle. Yhteistoiminnalliset järjestelmät ovat hajautettuna useisiin eri kantoihin ja järjestelmiin eivätkä ne ole loogisesti yksi järjestelmä. Niiden käyttösovellukset osaavat kuitenkin yhdistää järjestelmät ja muuttaa dataa sopivammaksi. Verkkojärjestelmät hyödyntävät verkkoteknologioita ja toimivat asiakas-palvelin-mallin mukaan. P2P-järjestelmät koostuvat identtisistä solmuista, jotka jakavat dataa ja palveluita ja täten täyttävät yhdessä käyttäjien tarpeet. (Batini et al. 2009, s. 11.) Näiden lisäksi voidaan nykyään myös ajatella olevan kokonaisvaltaisia järjestelmiä, jotka tarjoavat niin tietokannan datan säilyttämiseksi kuin myös käyttäjäsovelluksen dataan pääsemiseksi ja sen muokkaamiseksi.

## 2.3 Laadun arvioinnin menetelmät

Tässä luvussa tarkastellaan datan laadun eri arviointimenetelmiä. Myös menetelmien etuja ja haittoja sekä toteutuksen helppoutta arvioidaan. Luvussa käsitellään AIMQ-menetelmä, TDQM-menetelmä sekä IPMAP-työkalu. Muita, tässä työssä tarkastelemattomia menetelmiä ovat muun muassa Goal-Question-Metric-malli (Bobrowski et al. 1999), DQAF-malli (Sebastian-Coleman 2013) sekä Comprehensive Data Quality (Batini et al. 2006).

### 2.3.1 AIMQ-menetelmä

Lee et al. (2002) ovat luoneet AIMQ-menetelmän (AIM quality) datan laadun arvioinnille. Menetelmä tarjoaa viitekehysten datan laadulle, kyselylomakkeen datan laadun arviointiin sekä kaksi analyysimenetelmää saatujen tulosten arvioimiseksi. Datan laadun viitekehys perustuu PSP/IQ-malliin, ja se koostuu neljästä solusta: laadukkuus, luotettavuus, hyödyllisyys ja käytettävyys. Datan laadun eri ulottuvuudet jaotellaan näihin neljään soluun, ja ne vastaavat kehyksen solujen toteutumisesta. Taulukossa 4 on esitetty AIMQ-menetelmässä käytetty datan laadun viitekehys.

*Taulukko 4. AIMQ-menetelmän datan laadun viitekehys (Lee et al. 2002, s. 137)*

	Määrittelyjen mukaisuus	Kohtaa kuluttajien vaatimukset
<b>Tuotteen laatu</b>	Laadukas data Oikeellisuus Tiivis esitystapa Täydellisyys Johdonmukainen esitystapa	Hyödyllinen data Sopiva määrä Relevanssi Ymmärrettävyys Tulkittavuus Objektiivisuus
<b>Palvelun laatu</b>	Luotettava data Oikea-aikaisuus Tietoturvallisuus	Käytettävä data Uskottavuus Saatavuus Helppokäyttöisyys Maine

Viitekehyksessä datan laatua lähestytään kahdesta näkökulmasta: tuotteen laatu ja palvelun laatu. Data voidaan nähdä tuotteena, sillä datan tuottaminen vastaa tuotantoprosessia, jossa lopputuote on tietokantaan talletettu data (Ballou et al. 1998). Toisaalta data voidaan nähdä palveluna, sillä siihen liittyy toimintoja myös tallettamisen jälkeenkin: dataa haetaan ja käytetään jatkuvasti (Kahn et al. 2002, s. 186).

Kyselylomakkeella mitataan datan laatu jokaisen ulottuvuuden kohdalla, minkä jälkeen tuloksista lasketaan keskiarvo jokaiselle neljälle solulle. Jokaiselle ulottuvuudelle on 4-5 kohtaa, ja kyselylomakkeessa on yhteensä 65 kohtaa (Lee et al. 2002, s. 138). Kyselylomakkeen kohdat ovat väittämiä liittyen tarkasteltavaan laadun ulottuvuuteen. Liitteessä B on esitetty AIMQ-kyselylomakkeen kohdat.

AIMQ sisältää kaksi analyysimenetelmää tulosten arvioimiseksi. Toinen menetelmä on benchmarking-analyysi, jossa saatuja tuloksia verrataan esimerkiksi menestyvien kilpailijoiden tai toimialaa johtavien yritysten tuloksiin. Toinen menetelmä on kuiluanalyysi. Tässä tuloksia vertaillaan eri roolien kesken: esimerkiksi IT-ammattilaiset (datan ylläpitäjät), datan käyttäjät, datan tuottajat. Tämän menetelmän avulla voidaan arvioida, näkevätkö eri roolit datan laadun eri tavalla. Jos esimerkiksi datan ylläpitäjät näkevät datan

laadun parempana kuin loppukäyttäjät, eivät ylläpitäjät ole välttämättä tietoisia datan laadun ongelmista. (Lee et al. 2002, s. 141.) Batini et al. (2009, s. 39) huomauttavat, että kirjallisuus ei kuvaa benchmarking-tietokantaa, jota vaaditaan benchmarking-analyysin tekemiseen. Tämän vuoksi kannatettu lähestyminen tulosten arviointiin on kuiluanalyysi. Lisäksi case-esimerkkejä AIMQ-menetelmästä on olemassa vähän; esimerkiksi Helfert et al. (2009) sekä Moges (2014) ovat käyttäneet AIMQ-menetelmää kyselylomakkeen laatimisessa.

AIMQ-menetelmä soveltuu hyvin niin organisaation sisäiseen tarkasteluun kuin ulkopuolisten tarkastelijoiden tekemään datan laadun arviointiin. Se on melko helppo soveltaa kohdeyrityksen kontekstiin ja se on myös kevyt toteuttaa. Tulosten analyysi voi kuitenkin jäädä pinnalliseksi, jos vastaajia ei ole riittävästi tai he edustavat samaa roolia. Tällöin kuiluanalyysin tekeminen vaikeutuu.

Menetelmä voidaan myös nähdä hyvin subjektiivisena, sillä kun käytetään menetelmän tarjoamaa kyselylomaketta ja haastatellaan yrityksen työntekijöitä, saadaan hyvin subjektiivisia tuloksia. Batini et al. (2009, s. 17) mukaan AIMQ sisältääkin vain subjektiivisia mittareita, mutta benchmarking-analyysi tulosten arviointiin on kuitenkin objektiivinen. Koska datan laatu itsessään on subjektiivinen asia, ei objektiivisten mittareiden puuttuminen ole ongelma. Lisäksi datan loppukäyttäjien osallistaminen arviointiprosessiin on tärkeää, jolloin on käytettävä subjektiivisia mittareita.

### 2.3.2 Total Data Quality Management

Kasvattaakseen tuottavuutta yritysten tulee hallita tietoa samalla tavalla kuin tuotteita (Wang 1998, s. 58). Tämä tarkoittaa, että tietoa ja dataa tulee lähestyä tuoteajattelun näkökulmasta. Wang (1998) ehdottaa TDQM-menetelmää (Total Data Quality Management) datan laadun ongelmiin puuttumiseksi. Sen tarkoituksena on tuottaa korkealaatuisia *tietotuotteita* tiedon kuluttajille. TDQM perustuu tuotteiden valmistuksen alalla syntyneeseen Total Quality Management -ajatteluun (TQM). Taulukossa 5 havainnollistetaan tuotteiden ja tiedon valmistuksen samankaltaisuutta.

**Taulukko 5.** Tiedon valmistus verrattuna tuotteiden valmistukseen (Wang 1998, s. 59)

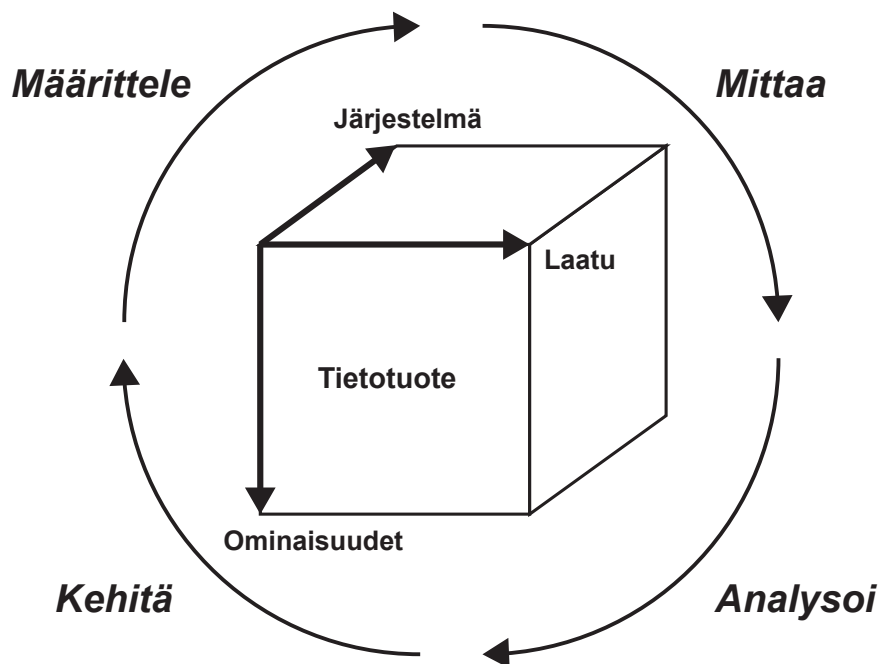
	Tuotteiden valmistus	Tiedon valmistus
<b>Syöte</b>	Raakamateriaali	Raakadata
<b>Prosessi</b>	Tuotantolinja	Tietojärjestelmä
<b>Lopputuote</b>	Fyysinen tuote	Tietotuote

Tuotteiden valmistuksessa raakamateriaalista valmistetaan tuotantolinjoilla fyysisiä tuotteita. Tiedon valmistuksessa puolestaan raakadatasta jalostetaan tietojärjestelmissä tietotuotteita. Tämä analogia toimii perusteluna TQM-ajattelun soveltamiselle datan laadun

arviointiin ja kehittämiseen (Wang 1998, s. 59). Batini et al. (2009, s. 35) mukaan TDQM:n tavoitteena on poistaa eroavuudet prosessien tuotteiden ja niiden käyttäjien vaatimusten välillä.

TDQM-menetelmässä tunnistetaan neljä eri käyttäjäryhmää. *Tiedon toimittajat* luovat ja keräävät dataa tietotuotetta varten. *Tiedon tuottajat* suunnittelevat, kehittävät tai ylläpitävät tietotuotteen tarvitsemaan dataa ja järjestelmäarkkitehtuuria. *Tiedon kuluttajat* käyttävät tietotuotetta työssään. *Tietotuotteen hallitsijat* vastaavat koko tietotuotteen tuotantoprosessin hallitsemisesta koko elinkaaren ajan. (Wang 1998, s. 60.) Eri roolit on tunnistettu tietotuoteajattelu huomioiden. Eri roolit ovat tekemisissä tiedon kanssa sen tuotantoprosessin eri vaiheissa. Tietotuotteen hallitsijat ohjaavat kuitenkin koko prosessia. Tietotuotteen käsite korostaa sitä, että tiedontuotantojärjestelmän lopputuotteella on arvo, joka voidaan välittää kuluttajalle (Wang 1998, s. 60).

TDQM koostuu neljästä vaiheesta: määrittele, mittaa, analysoi ja kehitä. Tämä iteratiivinen prosessi taas vaikuttaa sen keskellä olevaan tietotuotteeseen, jolla on laatulementti sekä erityispiirteet/ominaisuudet. Tietotuote kulkeutuu lisäksi tiedontuotantojärjestelmän (*Information Manufacturing System*) läpi. Kuvassa 2 on esitetty TDQM-menetelmän kaaviomalli.



**Kuva 2.** TDQM-menetelmän kaaviomalli (mukaillen Wang 1998, s. 60)

Määrittelyvaiheessa organisaatio määrittelee tietotuotteelle ominaisuudet, tiedon laadun vaatimukset sekä tiedontuotantojärjestelmän. Tietotuotteen ominaisuuksien korkean tason määrittelyssä tietotuote käsitteellistetään sen käyttäjille tarjoamien toimintojen kautta. Alemmalla tasolla määritellään tietotuotteen komponentit ja niiden väliset suhteet. Tiedon laadun vaatimusten määrittelyssä tunnistetaan vaatimukset eri roolien näkökul-

masta (toimittajat, tuottajat, kuluttajat ja hallitsijat). Vaatimukset voidaan vuorostaan jaloistaa tarkemmiksi objektiivisiksi mittareiksi, joita voidaan käyttää tiedon laadun arviointiin tai seurantaan. Tiedontuotantojärjestelmän määrittelyssä mallinnetaan prosessia, jonka kautta tieto kulkee toimittajalta kuluttajalle. (Wang 1998, s. 61-64.) Prosessin mallinnukseen voidaan käyttää IPMAP-työkalua (*Information Product Map*), jonka Shankaranarayanan et al. (2000) esittelevät tarkemmin. Shankaranarayanan et al. (2003) ja Shankaranarayanan & Cai (2006) ovat edelleen kehittäneet ja soveltaneet IPMAP-työkalua. Työkalua käsitellään tarkemmin luvussa 2.3.3. Prosessin mallinnukseen jälkeen on sitä mahdollista kehittää tehokkaammaksi, jolloin myös tietotuotteiden laatu paranee.

Mittausvaiheessa määritellään datan laadun mittarit. Mittarit voidaan ottaa käyttöön uuden tiedontuotantoprosessin yhteydessä tai liittää rutiineiksi olemassa olevaan prosessiin, jolloin niiden avulla saadaan tuloksia tiedon laadusta. (Wang 1998, s. 64.) Näitä tuloksia voidaan taas käyttää analyyseihin ja tuotantoprosessin edelleen kehittämiseen.

Analysointivaiheessa selvitetään mittaustulosten avulla nykyisten laatuongelmien juurisyyt. Analysoinnissa voidaan hyödyntää eri työkaluja, muun muassa statistista prosessin hallintaa (Statistical Process Control) Myös valittujen mittareiden sopivuutta voidaan arvioida. (Wang 1998, s. 64.) Tässä vaiheessa on myös hyödyllistä määritellä kehitystavoitteet menetelmän seuraavaa vaihetta varten.

Kehitysvaiheessa tulee tunnistaa tärkeimmät kehityksen osa-alueet. Näitä voivat olla esimerkiksi tietovirtojen ja työskentelyn yhteensovittaminen vastaavien tiedontuotantojärjestelmien kanssa sekä tietotuotteen avainominaisuuksien uudelleen määrittely liiketoimintatarpeiden mukaan. (Wang 1998, s. 65.) TDQM-menetelmässä on huomioitavaa, että prosessi on jatkuvaa. Toisin sanoen kehitysvaiheesta siirrytään uuteen määrittelyvaiheeseen, jolloin tavoitteena on tietotuotteen jatkuva kehittäminen.

Batini et al. (2009, s. 35) mukaan TDQM täyttää kaikki datan laadun arviointiprosessin vaiheet. Tietotuotteen ominaisuuksien määrittely vastaa data-analyysiä, tiedon laadun vaatimusten määrittely vastaa laatuvaatimusanalyysiä, tiedontuotantojärjestelmän mallinnus vastaa prosessimallinnusta sekä mittareiden määrittely ja soveltaminen vastaa laadun mittausta. Lisäksi TDQM sisältää juurisyyden etsimisen sekä avainosa-alueiden tunnistamisen kehitystä varten. Muun muassa Kovac et al. (1997) ja Kovac & Weickert (2002) ovat soveltaneet TDQM-menetelmää.

### 2.3.3 IPMAP

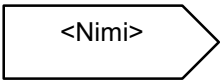

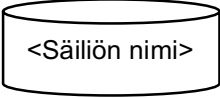
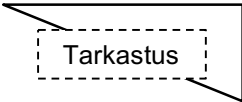
IPMAP (*Information Product Map*) on työkalu tietotuotteiden valmistuksen systemaattiseen mallintamiseen. Sen avulla organisaatio voi muun muassa visualisoida tietotuotteen valmistamista, ottaa käyttöön jatkuvan kehittämisen ja quality-at-source-laadunhallinnan sekä mitata tietotuotteen laatua. (Shankaranarayanan et al. 2000, s. 1.) IPMAP-työkalua voidaan käyttää yhdessä muiden laadun arvioinnin menetelmien kanssa. Erityisesti sitä

suositellaan käytettävän yhdessä TDQM-menetelmän tiedontuotantojärjestelmän mallinnuksen yhteydessä. IPMAP on laajennettu ja tarkemmin määritelty evoluutio Ballou et al. (1998) esittelemästä tietotuotteiden mallinnustyökalusta. Shankaranarayanan et al. (2003, s. 16-17) mukaan tämä aiemmin esitelty työkalu on kuitenkin tarkoitettu lopullisen tuotteen laadun ymmärtämiseen, ei tietotuotteen valmistusprosessin esittämiseen tai ymmärtämiseen. Shankaranarayanan et al. (2000, s. 2) kuvailevat IPMAP-työkalun esitystavan useita etuja:

- tietotuotteen hallitsija voi visualisoida valmistusprosessin kriittiset kohdat, jotka vaikuttavat tietotuotteen laatuun
- esitystapa mahdollistaa pullonkaulojen löytämisen valmistusprosessista sekä toimitusaikojen arvioimisen
- työkalu mahdollistaa osaprocessien vastuuhenkilöiden/omistajien tunnistamisen sekä quality-at-source-laadunhallinnan käyttöönoton
- mallinnus auttaa tietotuotteen hallitsijoita ymmärtämään valmistusprosessiin liittyvien organisaation yksiköiden sekä tietojärjestelmien rajoja
- työkalu mahdollistaa tietotuotteen laadun arvioinnin valmistusprosessin eri vaiheissa.

IPMAP-työkalun notaatiossa käytetään seitsemää erilaista symbolia eikä malleissa tule käyttää muita symboleita. Jokaisella symbolilla on jokin arvo, ja jokaiseen symboliin liittyy joukko attribuutteja. Symbolit liitetään toisiinsa nuolien avulla. Nuolen suunta kuvaa raakadatan tai tietotuotteen suuntaa valmistusprosessissa. Taulukossa 6 on esitetty IMAP-työkalun notaatio.

**Taulukko 6.** IPMAP-työkalun notaatio (Shankaranarayanan & Cai 2006, s. 306)

Symboli	Kuvaus
	<b>Datan lähde:</b> esittää jokaisen raakadatan lähdetä, joka tulee olla saatavissa, jotta tietotuote voidaan tuottaa.
	<b>Prosessointilohko:</b> esittää kaikkea raakadataan kohdistuvaa käsittelyä, laskentaa tai yhdistämistä, joita tarvitaan tietotuotteen tuottamiseen.
	<b>Datan säiliö:</b> säiliölohkota voidaan käyttää esittämään dataa, joka odottaa prosessointia tai joka talletetaan johonkin pitkäaikaisvarastoon.
	<b>Tarkastuslohko:</b> esittää datan laadun tarkastusta niille dattoille, jotka ovat tärkeitä virheettömän tietotuotteen tuottamisessa. Lohko ottaa syötteenä raakadataa ja tuottaa kaksi datavirtaa: virheetön data ja virheellinen data.

<div data-bbox="336 219 627 315" style="border: 1px solid black; padding: 5px; text-align: center;">         &lt;Nykyiset/uudet järjestelmien nimet&gt;       </div>	<b>Tietojärjestelmän raja:</b> käytetään, kun data kulkee järjestelmästä toiseen (esimerkiksi paperilta tietojärjestelmään tai päinvastoin). Lohko kuvaa muutosta raakadataan kun se kulkeutuu järjestelmästä toiseen.
<div data-bbox="344 371 619 472" style="border: 1px solid black; padding: 5px; text-align: center;">         ( &lt;Nykyiset/uudet yksiköiden nimet&gt; )       </div>	<b>Organisaation yksikön raja:</b> esittää tilanteita, joissa raakadata luovutetaan organisaation yksiköstä toiseen. Sitä käytetään määrittelemään tietotuotteen liikkumista yksiköiden tai organisaatioiden rajojen yli. Tämän lohkon tarkoitus on korostaa mahdollisia ongelmia, jotka syntyvät, kun tietotuote ylittää liiketoimintayksiköiden rajoja, ja vastuu tietotuotteesta muuttuu.
<div data-bbox="373 551 592 629" style="border: 1px solid black; padding: 5px; text-align: center;"> <div style="display: inline-block; transform: rotate(-45deg); width: 0; height: 0; border-left: 10px solid transparent; border-right: 10px solid transparent; border-bottom: 15px solid black;"></div>         &lt;Nimi&gt;       </div>	<b>Datan kuluttaja:</b> kuluttaja käyttää tätä lohkoa esittämään dataa, joka vastaa lopullista, valmista tietotuotetta. Tähän lohkoon liittyy tietotuotteesta vastuussa oleva liiketoimintayksikkö, tietotuotteen varsinainen käyttäjä sekä kokoelma eri datoja, jotka yhdessä muodostavat tietotuotteen.

Tärkeänä tavoitteena IPMAP-esitystä kehitettäessä on kyky sisällyttää datan laadun ulottuvuuksia jokaiseen lohkoon (Shankaranarayanan et al. 2000, s. 15). IPMAP voi auttaa organisaatiota hallitsemaan datan laatua ja ottamaan TDQM-lähestymisen datan laatuun (Shankaranarayanan et al. 2003, s. 1). IPMAP myös mahdollistaa datan laadun arvioinnin sekä objektiivisesti että kontekstisidonnaisesti (Shankaranarayanan & Cai 2006, s. 1). Kirjallisuudesta löytyy kuitenkin vähän case-esimerkkejä IPMAP-työkalun soveltamisesta. Kuitenkin muun muassa Davidson et al. (2004) ovat käyttäneet sovellettua versiota IPMAP-työkalusta. Työkaluun liittyviä laskentakaavoja, joilla laatu voidaan esittää kvantitatiivisesti valmistusprosessin eri vaiheissa, ei ole esitetty tässä työssä. Esimerkiksi Shankaranarayanan & Cai (2006) esittelevät työkaluun liittyvät kaavat yksityiskohtaisesti.

## 2.4 Yhteenveto

Datan laadun arvioinnin menetelmät mahdollistavat datan laadun seurannan. Näitä menetelmiä käyttämällä yritykset voivat joko seurata systemaattisesti omaa suoriutumistaan tai suorittaa kertaluonteisia arviointoja, joilla kartoitetaan datan laadun nykytilanne. Seuraamalla ja arvioimalla datan laatua yritykset voivat ottaa kantaa datalähteisiin tai dataa käsitteleviin prosesseihin, jotka tuottavat heikkolaatuista dataa tai heikentävät datan laatua.

Menetelmiä datan laadun arviointiin on monia, ja yritysten tulee valita itselleen parhaiten sopiva menetelmä tai vaihtoehtoisesti yhdistellä aspekteja eri menetelmistä. AIMQ-menetelmä on kevyt toteuttaa, ja se mahdollistaa subjektiivisten näkemysten keräämisen. TDQM on taas kattava menetelmä, joka korostaa tietotuotteen käsitteen merkitystä. Osa TDQM-menetelmästä on prosessimallinnus, jossa kuvataan, miten tieto kulkee toimittajilta kuluttajille. Datavirtoja ja tuottoprosessin eri vaiheita voidaan kuvata esimerkiksi IPMAP-työkalulla. Se auttaa muun muassa datan käyttäjiä mallintamaan ja ymmärtämään datan tuottoprosessia sekä arvioimaan lopullisen tietotuotteen sekä tietolähteen laatua.



Arvioidessaan eri menetelmien soveltuvuutta omaan toimintaympäristöön, yritykset voivat hyödyntää datan laadun arvioinnin menetelmien vertailukohtia. Jotkin menetelmät käyttävät objektiivisia mittareita, toiset subjektiivisia mittareita. Jotkin menetelmät saattavat käyttää molempia. Objektiiviset mittarit mahdollistavat datan laadun vertaamisen johonkin standardiin tai itseensä ajan suhteen. Subjektiiviset mittarit taas mahdollistavat esimerkiksi datan käyttäjien näkemysten huomioimisen. Lisäksi yritysten on valittava sopiva strategia, jos arviointimenetelmä sisältää datan laadun kehittämisen. Datalähtöinen strategia ottaa kantaa olemassa olevan datan laatuun, kun taas prosessilähtöisellä strategialla voidaan varmistaa jatkossakin tuotettavan datan hyvä laatu.

Datan laadun arvioinnissa on olennaista ymmärtää, mistä datan laatu koostuu. Datan laadun ulottuvuudet kuvaavat datan piirteitä, joita voidaan mitata. Laadun arvioinnissa yritysten tulee tunnistaa itselleen kriittisimmät ulottuvuudet ja keskittyä niihin arvioinnissaan.

### 3. DATAN HALLINNOINTI

Tämän luvun alkuun on syytä määritellä datan hallintaan liittyviä käsitteitä ja perustella tässä työssä käytettyjä suomennoksia. Englannin termi *data management* on käännetty datan hallinnaksi. Termi *data governance* on käännetty taas datan hallinnoinniksi vakiintuneen suomennoksen puuttuessa. Nämä käsitteet eivät ole synonyymejä keskenään. Datan hallinnointi viittaa siihen, mitä päätöksiä pitää tehdä IT:n tehokkaan hallitsemisen ja käyttämisen varmistamiseksi sekä kuka tekee nämä päätökset (esimerkiksi päättäntävalta datan standardeista); hallinta puolestaan sisältää päätöksien tekemisen ja niiden implementoinnin (esimerkiksi laadun mittareiden määrittely) (Khatri & Brown 2010, s. 1). Toisen näkemyksen mukaan datan hallinnointi tarjoaa viitekehyksen johtamiselle ja päätöksenteolle, kun taas datan hallinta on jokapäiväistä päätöksentekoa (Wende 2007, s. 419). Datan hallinnoinnissa ei siis ole kyse itse datasta, vaan liiketoimintaprosesseista, päätöksistä sekä sidosryhmien vuorovaikutuksesta (Karel 2014a).

Datan hallinnoinnille ei ole muodostunut yhtä vakiintunutta määritelmää. Kirjallisuudessa on kuitenkin esitetty useita, toisiaan täydentäviä määritelmiä:

*Datan hallinnointi viittaa siihen, kenellä organisaatiossa on päätöksentekovalta ja vastuu organisaation dataan liittyvästä päätöksenteosta (Khatri & Brown 2010, s. 149).*

*Datan hallinnointi on kokoelma päättäntävaltoja, prosesseja, standardeja, käytäntöjä ja teknologioita, joita tarvitaan tiedon hallitsemiseen, ylläpitämiseen ja hyödyntämiseen (Newman & Logan 2006, s. 3).*

*Datan hallinnointi liittyy organisaation osiin, sääntöihin, päättäntävaltoihin sekä ihmisten ja tietojärjestelmien vastuullisuuteen tietoprosesseissa. Lisäksi datan hallinnointi asettaa säännöt, joita johdon tulee noudattaa, kun organisaatio käyttää dataa. (Thomas 2006; Cheong & Chang 2007 mukaan.)*

*Datan hallinnointi on käytäntöjen ja prosessien määrittelyä ennakoivan ja tehokkaan datan hallinnan varmistamiseksi (Cheong & Chang 2007, s. 1001).*

*Datan hallinnointi on prosessi, jonka kautta yritys hallitsee datan määrää, yhdenmukaisuutta, käytettävyyttä, turvallisuutta sekä saatavuutta (Cohen 2006).*

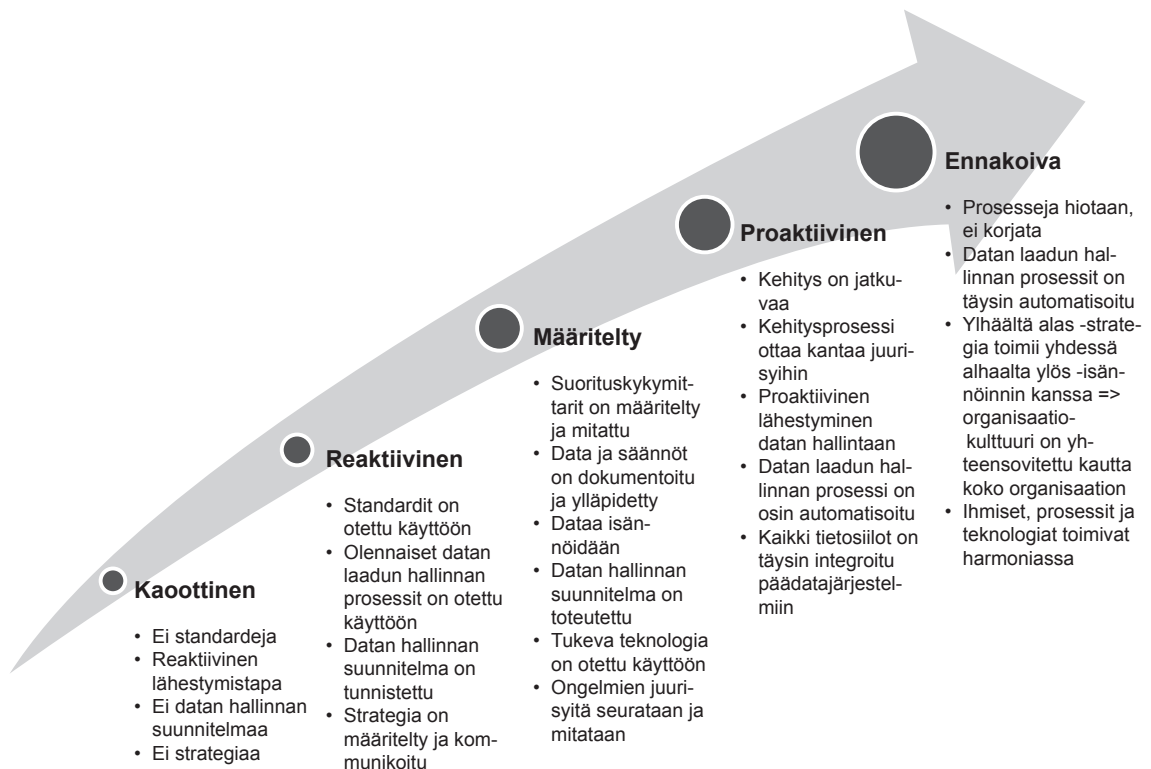
Eri määritelmillä on yhteistä se, että ne korostavat päätöksentekovallan ja vastuiden osoittamista sekä käytäntöjen ja prosessien merkitystä. Ihmiset ja prosessit ohjaavat dataa, minkä vuoksi datan hallinnointi on sekä ihmisten että teknologioiden hallintaa (Thomas 2006; Cheong & Chang 2007 mukaan). Lisäksi datan laadun kehittämisessä onnistuminen vaatii holistisen lähestymistavan, jossa tulee keskittyä ihmisiin, prosesseihin ja teknologiaan (Friedman 2006, s. 4). Dataa tulee siis hallinnoida, jos yritykset haluavat ottaa kantaa datan laadun epäkohtiin (Cheong & Chang 2007, s. 4). Näiden näkemysten mukaan

hallinnointi on datan laadun kehittämishankkeille tärkeä ja laajempi kokonaisuus, joka ottaa teknologisten näkökulmien lisäksi kantaa myös organisatorisiin kysymyksiin.

Datan hallinnointi voi parantaa yrityksen datan laatua, saatavuutta ja eheyttä kannustamalla yritystä jäsennettyyn käytäntöjen määrittelyyn sekä organisaatioiden väliseen yhteistyöhön (IBM 2007, s. 3). Datan hallinnoinnin hyötyjä ovat muun muassa laadukkaampi ja nopeampi päätöksenteko, alentuneet kustannukset, kehittyneempi liiketoimintatiedon raportointi sekä tehokkaampi viranomaissäätelyiden noudattaminen (Waddington 2010, s. 15). Lisäksi datan hallinnoinnilla voidaan kehittää yrityksen toimintaa läpinäkyvämmäksi, parantaa tietovirtoja sekä muuttaa yrityksen käytäntöjä johdonmukaisemmiksi (Newman & Logan 2006, s. 4). Muita datan hallinnoinnin ajureita ovat parantunut datan laatu ja tietoturvallisuus (Plotkin 2014, s. 7).

### 3.1 Hallinnoinnin tasot

Jotta yrityksessä voidaan lähteä kehittämään datan hallinnointia, on hyvä tietää mikä on yrityksen lähtötaso. Datan hallinnoinnin tehokkuutta jossain tietyssä organisaatiossa voidaan arvioida datan hallinnoinnin kypsyystason malleilla. Kypsyydellä viitataan tässä tapauksessa siihen, miten systemaattista hallinnointi on. Mitä tehokkaammin ja järjestelmällisemmin organisaatio hallinnoi dataansa, sitä kypsempää se on. Kuvassa 3 on esitetty datan hallinnoinnin kehityksen ja kypsyysen viisi eri tasoa.



**Kuva 3.** Datan hallinnoinnin tasot (mukaillen Brennan 2012)

Alkeellisin datan hallinnoinnin taso on *kaoottinen*. Tällä tasolla organisaatiolla ei ole käytössä mitään muodollisia käytäntöjä datan hallinnalle. Toinen taso on *reaktiivinen*, jossa on otettu käyttöön standardeja ja muodollisia prosesseja. Kolmas taso on *määritelty*, jossa organisaatio on tunnistanut suorituskymittarit, jotka mittaavat prosessien toimivuutta. Dataa myös isännöidään sen luomisesta kulutukseen ja käyttöön asti. Harvat yritykset ovat päässeet ohi määritellyn tason. *Proaktiiviseen* tasoon pääseminen vaatii jo huomattavaa keskittymistä ja investointia datan hallinnoinnin kehittämiseen. Kehityksen tulee olla jatkuvaa ja palaute johtaa prosessien kehittämiseen. Automatisoinnin taso on korkea, jotta inhimillisten virheiden määrä voidaan minimoida. Viimeinen taso on *ennakoiva*, jossa prosessit on automatisoitu ja ne ovat jäljitettäviä. Organisaatiossa edistetään vastuullisuutta ja omistajuutta. (Brennan 2012.) Realistinen tavoite useille yrityksille on määritelty taso. On kuitenkin syytä huomata, että jo tällä tasolla datan hallinnointi on vahvalla pohjalla. Tämä taas heijastuu myönteisesti muun muassa datan laatuun ja käytäntöihin. Määriteltyä edistyneemmät tasot on vaikea saavuttaa ilman pitkäaikaista sitoutumista datan hallinnoinnin kehittämiseen.

Myös IBM (2007, s. 7) tarjoaa mallin yrityksen datan hallinnoinnin tason arvioinnille:

- **Taso 1:** Lain ja säännöstelyn vaatimat käytännöt on otettu käyttöön. Näille käytännöille kriittinen data on tunnistettu. Riskien arviointia voidaan myös tehdä kriittisen datan suojaamisesta.
- **Taso 2:** Enemmän dataan liittyviä säännöksiä on dokumentoitu ja julkaistu koko organisaatiolle. Ongelmia ratkaistaan enemmän proaktiivisella ja tiimiperusteisella lähestymistavalla sekä toistettavilla prosesseilla. Metadata on tärkeä osa kriittisten dataelementtien dokumentointia.
- **Taso 3:** Dataan liittyvät käytännöt ovat yksikäsitteisempiä ja selkeämpiä ja heijastavat organisaation dataan liittyviä periaatteita. Datan integraation mahdollisuuksia tunnistetaan ja hyödynnetään paremmin. Datan eheyden ja laadun riskien arviointi on osa organisaation projektimenettelyä.
- **Taso 4:** Organisaatio määrittelee datan arvon useammalle dataelementille ja asettaa arvoperusteisia käytäntöjä näiden päätösten pohjalta. Datan hallinnoinnin rakenteet kattavat koko organisaation. Datan hallinnoinnin menetelmät esitellään uusien projektien suunnitteluvaiheessa. Yrityksen datamallit dokumentoidaan ja julkaistaan.
- **Taso 5:** Datan hallinnointi on luonnollinen osa yrityksen toimintaa. Dataan liittyvien projektien tuottoastetta seurataan jatkuvasti. Innovointiin kannustetaan. Datan hallinnan liiketoiminta-arvo on tunnistettu ja datan hallinnan kustannuksia on helpompi hallita. Kustannukset pienenevät, kun prosesseista tulee entistä automatisoidumpia ja suoraviivaisempia.

Tämä malli noudattaa Brennanin mallin periaatteita, mutta eroaa yksityiskohdiltaan. Mitä korkeamman tason datan hallinnoinnista on kyse, sitä luonnollisempi osa yrityksen toi-

mintaa se on. Prosessien korkeaa automaatioastetta sekä proaktiivista ja ennakoivaa lähestymistä datan hallinnointiin korostetaan molemmissa malleissa. Loshinin (2011, s. 45) datan hallinnoinnin mallissa on myös viisi tasoa. Taulukossa 7 on esitetty kypsyysmalli datan hallinnoinnille.

**Taulukko 7.** *Datan hallinnoinnin kypsyysmalli (Loshin 2011, s. 49)*

Taso	Kuvaus
Varhainen	<ul style="list-style-type: none"> <li>Datan laadun hallintaa kommunikoidaan vähän tai ei ollenkaan</li> <li>IT on oletusarvo kaikille datan laadun kysymyksille</li> <li>Ei datan isännöintiä (<i>stewardship</i>)</li> <li>Vastuut datan korjaamisesta jaetaan ad hoc -tyyppisesti</li> </ul>
Toistettava	<ul style="list-style-type: none"> <li>Parhaat käytännöt kerätään ja jaetaan osallistujien kesken</li> <li>Avainhenkilöt yhteisöstä muodostavat työryhmän luodakseen ja suositellakseen datan hallinnoinnin ohjelman ja menettelytapoja</li> <li>Ohjaavia periaatteita ja datan laadun peruskirja muotoillaan</li> </ul>
Määritelty	<ul style="list-style-type: none"> <li>Organisaatorakenne datan hallinnoinnin valvonnalle on määritelty</li> <li>Ohjaavat periaatteet, peruskirja ja datan hallinnan menettelytavat on dokumentoitu</li> <li>Standardoitu näkemys datan isännöinnistä kautta organisaation, datan isännöintiohjelma on otettu käyttöön</li> <li>Operatiiviset datan hallinnoinnin menetelmät on määritelty</li> </ul>
Hallittu	<ul style="list-style-type: none"> <li>Datan hallinnoinnin komitea koostuen henkilöistä kautta organisaation on otettu käyttöön</li> <li>Yhteistyöhön perustuva datan hallinnoinnin komitea kokoontuu säännöllisesti</li> <li>Datan laadun palvelutasosopimukset ajavat operatiivista datan hallinnointia</li> <li>Tiimit jokaisesta osastosta käyttävät samanlaista hallinnoinnin kehystä sisäisesti</li> <li>Raportoinnin ja korjaamisen viitekehykset toimivat yhdessä soveltaen statistista prosessinhallintaa (Statistical Process Control) määritellyissä laadun rajoissa pysymiseksi</li> </ul>
Optimoitu	<ul style="list-style-type: none"> <li>Datan laadun suorituskykymittareita arvioidaan kehitysmahdollisuuksien löytämiseksi</li> <li>Työntekijöitä palkitaan datan hallinnoinnin suorituskykytavoitteiden saavuttamisesta</li> </ul>

Kyseisessä mallissa datan hallinnointi kehittyy alhaalta ylös, kun mahdollisuuksia tiedon jakamiselle annetaan, sekä ylhäältä alas, kun ihmisten ja päättävien elinten vastuut virallistuvat (Loshin 2011, s. 47). Mallissa myös korostetaan yhteistyön ja datan hallinnointiin keskittyvän työryhmän/komitean merkitystä. Lisäksi menettelytapojen, standardien ja suorituskykymittareiden määrittely ja dokumentointi on tärkeää.

Myös IT Governance Institutin COBIT 4.1 -julkaisussa (ITGI 2007) esitellään kypsyysmalli datan hallinnan kypsyydelle. COBIT on viitekehys, jota käytetään yrityksen tietohallinnon johtamiseen ja hallinnointiin. Se sisältää muun muassa parhaita käytäntöjä,

kypsyysmalleja sekä hallinnan tavoitteita. Viitekehystä voidaan käyttää tietohallinnon tavoitteiden ja prosessien yhteensovittamiseen liiketoiminnan tarpeiden kanssa. COBIT:n (ITGI 2007, s. 144) kypsyysmalli koostuu kuudesta tasosta:

- **0. Ei datan hallintaa:** Dataa ei nähdä yrityksen voimavarana. Datan omistajuutta ei ole osoitettu eikä vastuuta datan hallinnasta. Datan laatu ja tietoturva ovat heikkoja.
- **1. Varhainen/Ad hoc:** Organisaatio tunnistaa tarpeen tehokkaalle datan hallinnalle. Datan hallinta määritellään ad hoc -tyyppisesti. Vastuu datan hallinnasta ei ole selkeä eikä datan hallintaa kouluteta. Datan palautettavuus on varmistettu.
- **2. Toistettava mutta intuitiivinen:** Koko organisaatio tiedostaa tehokkaan datan hallinnan tarpeen. Korkean tason datan omistajuutta esiintyy. Tärkeimpiä datan hallinnan toimintoja seurataan (esimerkiksi varmuuskopiointi). Vastuu datan hallinnasta on epämuodollisesti osoitettu tietohallinnon avainhenkilöille.
- **3. Määritelty:** Datan hallinnan tarve niin tietohallinnossa kuin koko organisaatiossa ymmärretään kautta linjan. Vastuu datan hallinnasta on määritelty. Datan omistajuus on osoitettu vastuussa oleville henkilöille. Datan hallinnan prosessit on määritelty tietohallinnossa. Datan hallintaa seurataan jossain määrin. Alkeelliset suorituskykymittarit on määritelty. Datan hallinnan henkilöstöä koulutetaan.
- **4. Johdettu ja mitattava:** Datan hallinnan tarve ymmärretään ja vaaditut toiminnot hyväksytään koko organisaatiossa. Vastuut datan omistajuudesta ja hallinnasta on selkeästi määritelty, osoitettu ja kommunikoitu organisaatiossa. Prosessit ovat määriteltyjä, laajasti tunnettuja ja tietoa jaetaan. Tavoitteet ja suorituskykymittarit on määritelty ja niitä seurataan. Datan hallinnan henkilöstöä koulutetaan muodollisesti.
- **5. Optimoitu:** Tarve datan hallinnalle ja sen vaatimat toiminnot ymmärretään koko organisaatiossa. Tulevia tarpeita ja vaatimuksia tutkitaan proaktiivisesti. Vastuut datan omistajuudesta ja datan hallinnasta on selkeästi vakiinnutettu ja laajasti tiedossa koko organisaatiossa. Prosessit ovat muodollisia ja laajasti tiedossa, ja tiedon jakaminen on tavanomaista. Hienostuneita työkaluja käytetään ja datan hallinta on mahdollisimman automatisoitua. Tavoitteet ja suorituskykymittarit on linkitetty liiketoimintatavoitteisiin ja niitä seurataan. Mahdollisuuksia kehittää toimintaa tutkitaan jatkuvasti. Datan hallinnan henkilöstön koulutus on johdettua.

COBIT:n datan hallinnan kypsyysmalli ottaa kantaa myös organisaation asenteeseen datan hallintaa kohtaa. Mitä kypsempää organisaation datan hallinta on, sitä laajemmin ja syvemmin datan hallinnan merkitys ja datan hallintaan liittyvät toiminnot ymmärretään ja hyväksytään. Lisäksi se ottaa kantaa myös datan hallinnan kouluttamiseen ja siihen, kuinka selkeästi vastuut on määritelty ja osoitettu.

### 3.2 Hallinnoinnin roolit

Vastuullisuus ja strateginen osallistaminen ovat yksi kriittinen onnistumistekijä datan hallinnoimisessa (Marinos 2004). Organisatorisen rakenteen suunnittelu sekä roolien ja vastuiden suunnittelu voidaan nähdä datan hallinnan lanseeraamisen tärkeimpänä kysymyksenä (Panian 2010, s. 943). Lisäksi organisaation teknologia- ja liiketoimintaryhmien tulee työskennellä tiiviisti yhdessä, jotta yritys voi onnistua datan hallinnan hankkeessa (Malange et al. 2015, s. 1). Datan hallinnointi onkin liiketoiminnan ja IT:n yhteinen pyrkimys (Page 2011, s. 18). Tämän vuoksi datan tehokas hallinnoiminen vaatii selkeän rakenteen sekä organisatoriset roolit ja vastuut. Vastuunjaossa tulee myös ottaa huomioon sekä organisaation IT- että liiketoimintayksiköt. Kuvassa 4 on esitetty datan hallinnoinnin neljä organisatorista roolia Cohenin (2006) mukaan.

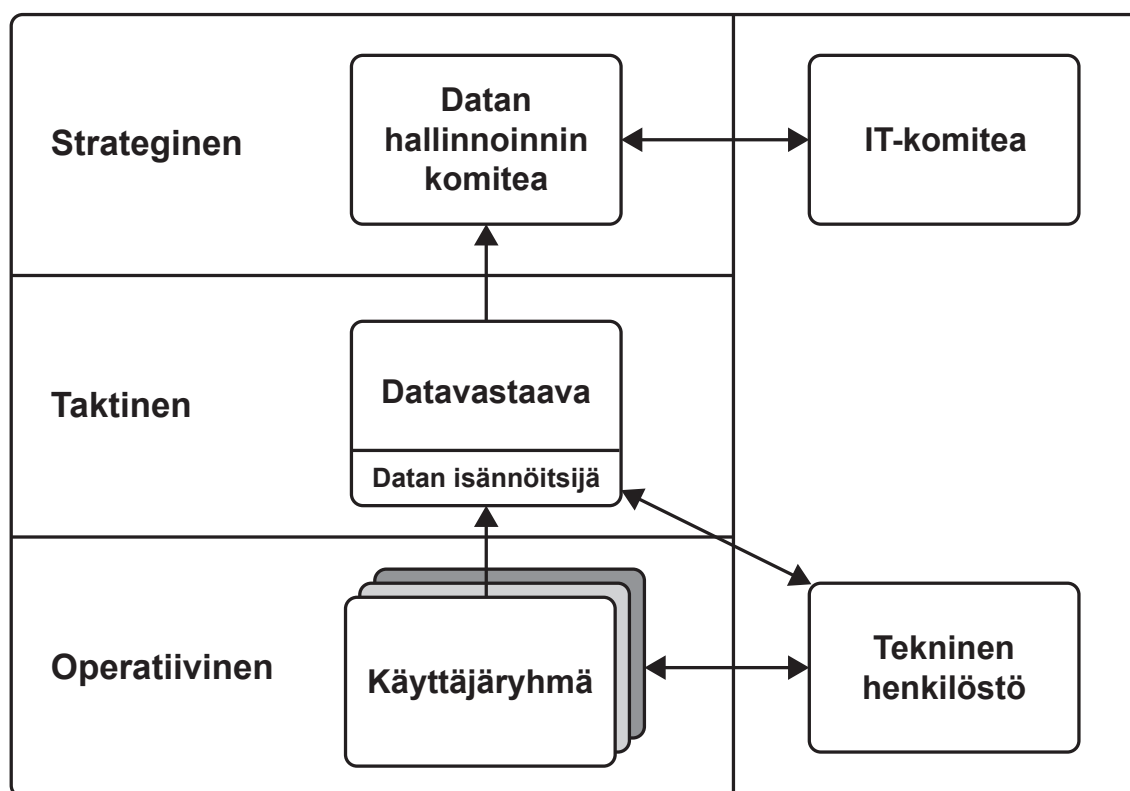


**Kuva 4.** Datan hallinnoinnin organisatoriset roolit (mukaillen Cohen 2006)

Usein kolme datan hallinnoinnin organisatorista roolia, isännöitsijä, omistaja ja hallitsija, nähdään käytännössä synonyymeina. Näin ei kuitenkaan ole. Roolien nimet eivät ole tärkeitä, mutta roolien vastuut ovat. *Isännöitsijät* määrittelevät datan hallinnan käytäntöjä ja neuvovat datan omistajia ja hallitsijoita niiden käyttöönotossa. *Omistajat* taas kantavat pääasiallisen vastuun tietovaatimusten määrittelystä. Datan *hallitsijat* toimivat yhdessä

isännöitsijöiden ja omistajien kanssa ottaakseen käyttöön datan hallinnan käytännöt. He vastaavat myös tiedon toimittamisesta muun organisaation käyttöön. Datan *käyttäjien* tulee ymmärtää tietotarpeensa riittävän hyvin ja noudattaa datan hallinnan käytäntöjä. He myös kommunikoivat datan hallitsijoiden kanssa sovelluksista ja teknologioista – mikä toimii ja mikä ei. Loppukäyttäjien tarpeet toimivat datan hallinnoinnin ajurina. (Cohen 2006.) Datan hallinnoinnin kolme muuta roolia palvelevat datan käyttäjien tarpeita. Käyttäjienkin on kuitenkin tiedettävä omat vastuunsa datan hallinnoinnissa, jotta se olisi tehokasta. Tärkeää on myös huomata, että datan isännöitsijä, omistaja ja hallitsija eivät ole yksi ja sama henkilö.

Myös Cheong & Chang (2007, s. 1005) esittävät mahdollisen datan hallinnoinnin rakenteen ja roolijaon. Rakenne koostuu datan hallinnoinnin komiteasta, IT-komiteasta, datan isännöitsijästä, käyttäjistä sekä teknisestä henkilöstöstä. Kuvassa 5 on esitetty datan hallinnoinnin rakenne.



**Kuva 5.** Datan hallinnoinnin rakenne (mukaillen Cheong & Chang 2007, s. 1005)

Datan hallinnoinnin komitea koostuu eri osastojen johtajista, joilla on kiinnostusta datavarojen hallintaan. Sen vastuu on toimintaohjeiden hyväksyminen, osastojen välisten ongelmien ratkominen, toimiminen yhdessä IT-komitean kanssa strategisella tasolla sekä IT- ja liiketoiminta-aloitteiden yhteensovittaminen. Datavastaava on vastuussa datavarojen laadusta sekä käyttäjryhmän kokouksissa esille nousseiden ongelmien ratkomisesta. Lisäksi datavastaava vastaa muutoshallinnasta, sidosryhmien hallinnasta, muuttaa strate-

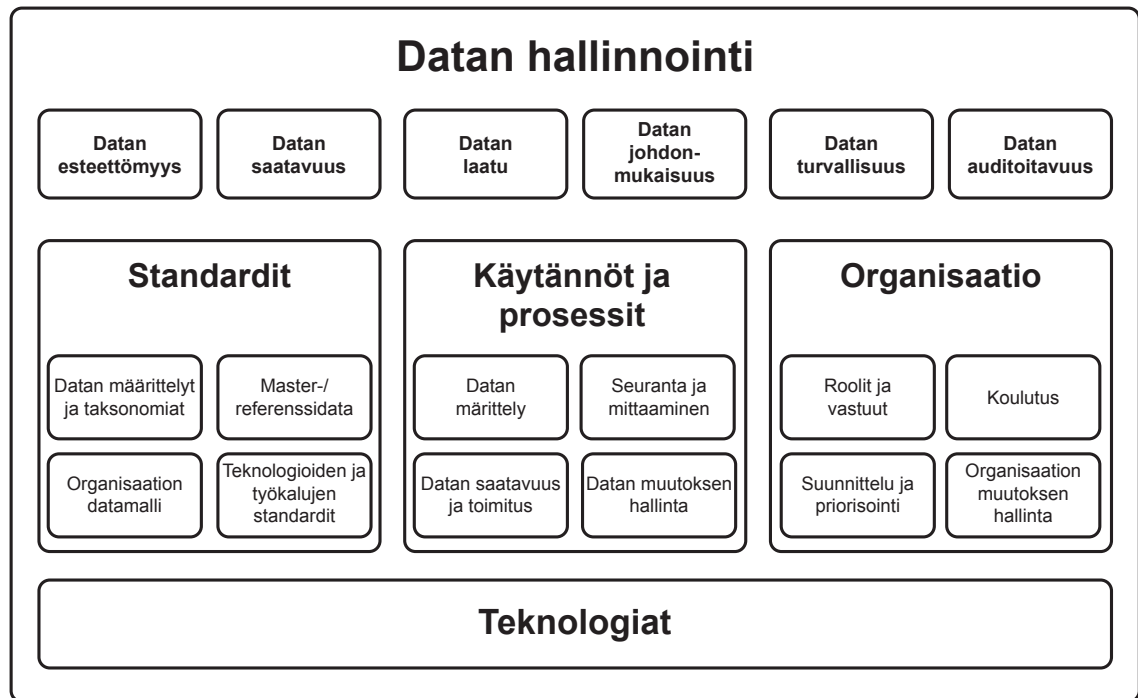


gisen suunnitelman taktiseksi suunnitelmaksi ja varmistaa, että data soveltuu tarkoitukseen. *Datan isännöitsijöillä* on yksityiskohtaista tietämystä liiketoimintaprosesseista ja datavaatimuksista. Lisäksi heillä on hyvää tietoteknistä tietämystä, jonka avulla he voivat muuttaa liiketoimintavaatimukset teknisiksi vaatimuksiksi. He toimivat datavastaavien alaisuudessa ja vastaavat taktisen suunnitelman toteuttamisesta. Isännöitsijät myös ohjaavat käyttäjäryhmien kokouksia ja kouluttavat käyttäjiä. Käyttäjäryhmien kokouksiin kutsutaan sidosryhmiä useista eri osastoista. Nämä avainsidosryhmät koostuvat henkilöistä, jotka keräävät, prosessoivat ja raportoivat dataa. *Tekninen IT-henkilöstö* myös kutsutaan kokouksiin, jotta heidän tekninen osaaminen on saatavissa kokouksessa. *Käyttäjien* vastuulla on dataan liittyvien ongelmien esiin tuominen, uusien toiminnallisuuksien toivominen sekä raportoinnin vaatimusten tarkentaminen. (Cheong & Chang 2007, s. 1005.)

Sekä Cohenin että Cheong & Changin mallissa korostuvat datan isännöinti, liiketoiminta- ja IT-yksiköiden yhteistyö sekä käyttäjien osallistaminen datan hallinnointiin. Cheong & Changin mallissa erotellaan kuitenkin strateginen, taktinen ja operatiivinen vastuu eri henkilöille/elimille. Cohenin malli ei ota kantaa datan hallinnoinnin strategian luomiseen ja sen jalkauttamiseen. Strategian muodostaminen on kuitenkin tärkeä osa datan hallinnoinnin prosessia (Adler 2007). Tämän vuoksi on syytä ottaa huomioon datan hallinnoinnin strategiset näkökulmat tukeutuessa Cohenin malliin.

### 3.3 Hallinnoinnin viitekehykset ja prosessi

Datan hallinnoinnin viitekehyksiä voidaan hyödyntää, kun ollaan kehittämässä tehokasta datan hallinnoinnin lähestymistapaa, strategiaa ja suunnitelmaa. Viitekehys on laaja linkitettyjen osien kuvaus, joka tukee jotain tiettyä lähestymistapaa tiettyyn tavoitteeseen ja toimii ohjeena, jota voidaan muokata tarvittaessa lisäämällä tai poistamalla osia (BusinessDictionary 2015b). Datan hallinnoinnin suunnittelu vaatii etääntymistä päivittäisestä päätöksenteosta ja keskittymistä tulevaisuudessa tehtävien päätösten tunnistamiseen sekä siihen, kenen niitä tulisi tehdä. (Khatrī & Brown 2010, s. 1.) Hallinnoinnin viitekehykset voivat olla hyödyksi, kun yritys ryhtyy datan hallinnoinnin hankkeeseen. Niiden avulla voidaan muun muassa tunnistaa olennaiset datan hallinnoinnin osa-alueet. Kuvassa 6 on esitetty datan hallinnoinnin viitekehys ja sen komponentit Panianin (2010, s. 944) mukaan.



**Kuva 6.** Datan hallinnoinnin viitekehyksen komponentit (mukaillen Panian 2010, s. 944)

Viitekehyksessä datalle on annettu kuusi tärkeää piirrettä: *esteettömyys* (dataan päästään käsiksi), *saatavuus* (datan on saatavilla, kun sitä tarvitaan), *laatu* (täydellisyys, tarkkuus, eheys), *johdonmukaisuus* (data on johdonmukaista kautta järjestelmien), *turvallisuus* (turvallinen pääsy dataan) ja *auditoidavuus* (jäljitettävyyys). Datan hallinnoinnin neljä komponenttia ovat standardit, käytännöt ja prosessit, organisaatio sekä teknologiat. *Standardit* sisältää datan määritelmät, taksonomiat, datamallit sekä tekniset dataan liittyvät standardit. *Käytännöt ja prosessit* kattaa dataan liittyvät liiketoimintasäännöt, dataan pääsyn ja sen toimituksen hallinnan, jatkuvan seurannan ja mittaamisen sekä datan muutosten hallinnan. *Organisaatio* sisältää roolien ja vastuiden määrittelyn, koulutuksen ja organisaation muutosten suunnittelun. *Teknologiat* sisältää muun muassa datan prosessointiin ja sen käsittelyn automatisointiin liittyvät järjestelmät ja työkalut. (Panian 2010, s. 943.)

Datan hallinnoinnin viitekehys voidaan ymmärtää seuraavalla tavalla. Teknologia toimii datan hallinnoinnin mahdollistajana. Standardit, käytännöt ja prosessit sekä organisaatiot linkittyvät teknologiaan ja hyödyntävät sitä. Näiden neljän ison komponentin toimivuus taas heijastuu datan laadun ulottuvuuksiin, jotka ovat viitekehyksessä ylhäällä. Datan laatu siis riippuu siitä, kuinka hyvin sen alla olevat datan hallinnoinnin komponentit toimivat.

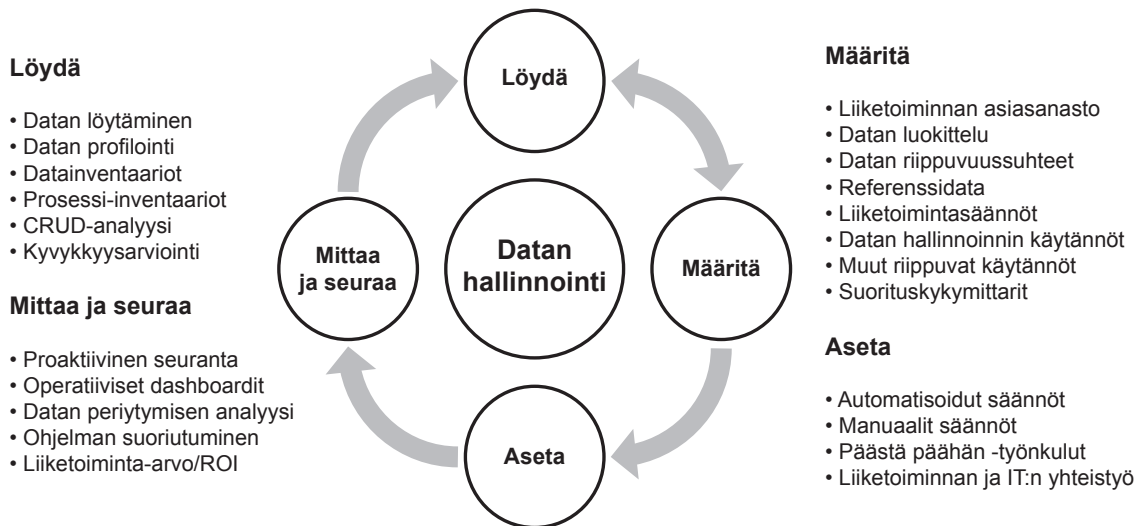
Myös Khatri & Brown (2010) ehdottavat datan hallinnoinnille viisi olennaista päätöksenteon piiriä: datan periaatteet, datan laatu, metadata, datan saatavuus ja datan elinkaari. Kukin piiri kattaa sille ominaiset päätökset sekä organisatoriset roolit ja vastuut. Taulukossa 8 on esitetty päätöksenteon piirit sekä niiden kuvaukset.

Datan hallinnoinnin suunnittelussa jokaisen päätöksenteon piirin vastuualue on jotain keskitetyn ja hajautetun välillä. Esimerkiksi keskitetty ryhmä yrityksen datan tuntevia johtajia voi suunnitella dataan liittyvät periaatteet koko organisaation käyttöön, kun taas datan laatuun liittyvä päätöksenteko voidaan hajauttaa datan eri omistajille eri liiketoimintayksikköihin. (Khatri & Brown 2010, s. 151-152.) Metadatalle, datan saatavuudelle ja datan elinkaarelle tulee myös pohtia, miten vastuut näistä jaetaan organisaatiossa. Yrityksen suunnitellessa datan hallinnointia, voidaan päätöksenteon piireistä ja niiden hajautuksesta tehdä taulukko. Tämän taulukon avulla on helppo nähdä, miten päätöksenteko on hajautettu kyseisessä organisaatiossa.

**Taulukko 8.** Datan hallinnoinnin päätöksenteon piirit ja kuvaukset (Khatri & Brown 2010, s. 149)

Datan hallinnoinnin päätöksenteon piiri	Päätettävät asiat	Vastuullinen rooli
<b>Datan periaatteet</b>  Datan roolin selkeyttäminen voimavarana	<ul style="list-style-type: none"> <li>Mikä on datan liiketoiminnallinen käyttö?</li> <li>Mitkä ovat toivotut menettelytavat datan käytölle?</li> <li>Miten sääntelyt vaikuttavat datan käyttämiseen?</li> </ul>	<ul style="list-style-type: none"> <li>Omistaja</li> <li>Isännöitsijä</li> <li>Tuottaja</li> <li>Käyttäjä</li> <li>Komitea</li> </ul>
<b>Datan laatu</b>  Datan laatuvaatimusten määrittely	<ul style="list-style-type: none"> <li>Mitkä ovat datan laadun standardit liittyen mm. tarkkuuteen, oikea-aikaisuuteen, täydellisyyteen ja uskottavuuteen?</li> <li>Miten datan laadusta kommunikoidaan?</li> <li>Miten datan laatua arvioidaan?</li> </ul>	<ul style="list-style-type: none"> <li>Omistaja</li> <li>Datan laatuvas- taava</li> <li>Datan laadun ar- vioija</li> </ul>
<b>Metadatan</b>  Datan sisällön ja se- mantiikan määrittely	<ul style="list-style-type: none"> <li>Miten data määritellään ja mallin-etaan niin, että se on jatkuvasti tulkitta- vissa?</li> <li>Miten erityyppiset metadatat pidetään ajan tasalla?</li> </ul>	<ul style="list-style-type: none"> <li>Data-arkkitehti</li> <li>Datan mallintaja</li> <li>Datan mallinnus- insinööri</li> </ul>
<b>Datan saatavuus</b>  Datan saatavuuden vaatimusten määrittely	<ul style="list-style-type: none"> <li>Mikä on datan liiketoiminta-arvo?</li> <li>Miten riskienarviointi suoritetaan jat- kuvasti?</li> <li>Mitkä ovat datan saatavuuteen liittyvät standardit?</li> <li>Miten tietoisuutta tietoturvallisuudesta levitetään?</li> <li>Miten datan varmuuskopiointi ja pa- lautus on varmistettu?</li> </ul>	<ul style="list-style-type: none"> <li>Omistaja</li> <li>Tietoturvallisuus- päällikkö</li> <li>Tietoturva-analyy- tikko</li> </ul>
<b>Datan elinkaari</b>  Datan määrittelyn, tuot- tamisen, säilyttämisen ja käytöstä poiston määrittäminen	<ul style="list-style-type: none"> <li>Miten datan inventaario suoritetaan?</li> <li>Miten datan elinkaari eri datatyypeille on määritelty?</li> <li>Miten viranomaisen sääntely vaikuttaa datan elinkaareen?</li> </ul>	<ul style="list-style-type: none"> <li>Data-arkkitehti</li> <li>Tietoketjuvas- taava</li> </ul>

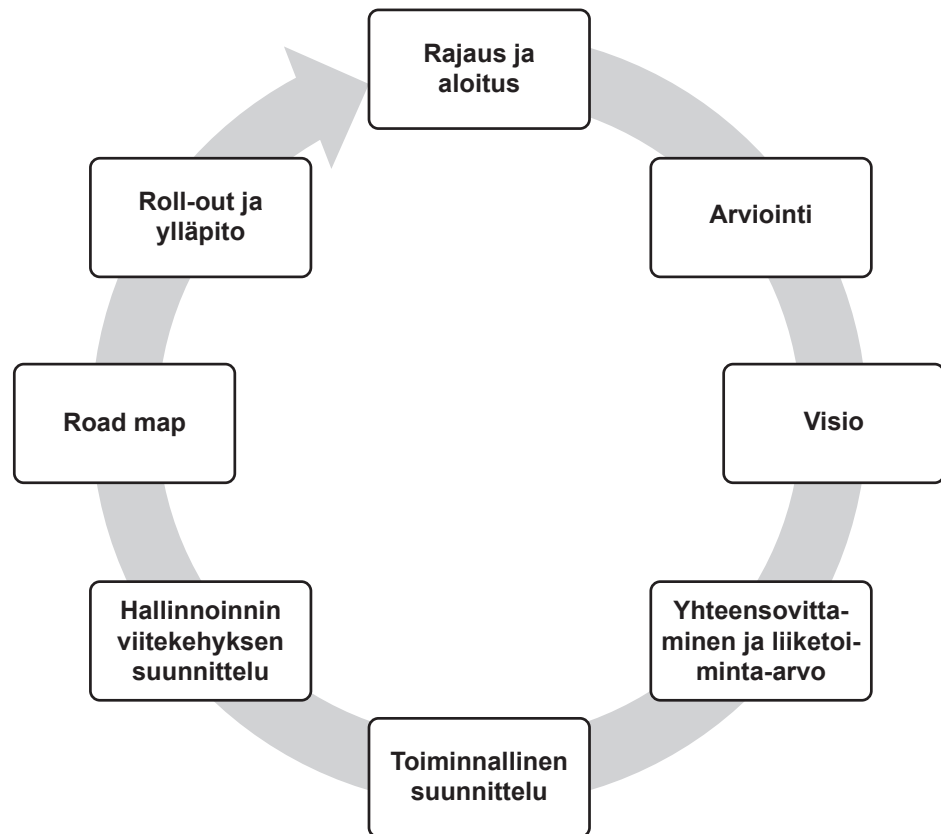
Jotta dataa voidaan tehokkaasti hallita voimavarana, tulee datan hallinnointia johtaa kuin muitakin liiketoimintoja. Samoin kuin muut liiketoiminnot, myös datan hallinnointi koostuu osaprosesseista. (Karel 2014b.) Prosessimainen datan hallinnoinnin viitekehys auttaa hahmottamaan hallinnoinnin eri vaiheet. Prosessimaisuus myös korostaa datan hallinnoinnin jatkuvuutta; kyse ei ole vain kertaluonteisesta hankkeesta. Kuvassa 7 on esitetty datan hallinnoinnin nelivaiheinen prosessimalli.



**Kuva 7.** Datan hallinnointi ja sen osaprosessit (mukaillen Karel 2014b)

Datan hallinnointi koostuu neljästä ydinprosessista, jotka taas koostuvat useasta osaprosessista. *Löydä*-vaiheessa tunnistetaan organisaation datan elinkaaren, riippuvien liiketoimintaprosessien, kyvykkyyksien sekä itse datan tämänhetkinen tilanne. Tämän vaiheen osaprosessien avulla voidaan organisaatiolle luoda datan hallinnoinnin strategia, prioriteetit, käytännöt, standardit, arkkitehtuuri sekä tulevaisuuden visio. Tämä vaihe toimii yhdessä *määritä*-vaiheen kanssa. Sen tarkoituksena on dokumentoida datan määritelmät, liiketoimintakonteksti, riippuvuussuhteet, käytännöt säännöt, standardit, prosessit sekä mittausstrategiat. Löytäminen ohjaa määrittelyä, ja määrittely puolestaan ohjaa kohdistetumpaa löytämistä. *Aseta*-vaiheessa prosessit pyrkivät ottamaan datan hallinnoinnin toiminnot käyttöön sekä saamaan organisaation noudattamaan datan hallinnoinnin käytäntöjä, liiketoimintasääntöjä, isännöintiprosesseja, työnkulkuja sekä annettuja vastuuta. *Mittaa ja seuraa* -vaiheessa mitataan datan hallinnoinnin tuottama lisäarvo, seurataan määriteltujen käytäntöjen noudattamista sekä varmistetaan toiminnan läpinäkyvyys. (Karel 2014b.) Prosessissa on huomioitavaa sen iteratiivisuus. Mittaa ja seuraa -vaiheen jälkeen siirrytään uudestaan löydä-vaiheeseen, jolloin prosessi alkaa uudelleen. Tavoitteena on datan hallinnoinnin jatkuva kehittäminen.

Myös Ladley (2012, s. 43) esittelee prosessin datan hallinnointiin. Tämä malli keskittyy enemmän datan hallinnoinnin perustamiseen, mutta mahdollistaa myös sen jatkuvan kehittämisen. Mallin avulla voidaan määritellä, suunnitella, jalkauttaa ja alkaa hallita datan hallinnoinnin hanketta. Kuvassa 8 on esitetty datan hallinnoinnin perustamisen prosessi.



**Kuva 8.** *Datan hallinnoinnin perustamisen prosessi (mukaillen Ladley 2012, s. 43)*

Prosessin kahdeksan vaihetta ohjaavat datan hallinnoinnin perustamista ja organisaation siirtämistä hallinnoituun tilaan:

#### *Rajausta ja aloitus*

Tässä vaiheessa tunnistetaan organisaatioyksiköt, jotka tulevat olemaan datan hallinnoinnin alaisia. Myös muut dataan liittyvät hankkeet tulee huomioida, esimerkiksi eri tietojärjestelmäprojektit. Jos vastuuroolien osoittaminen on uutta organisaatiolle, tulee henkilöstöhallinta sisällyttää hankkeen rajaukseen. Myös organisaation luonteella on merkitystä; jos vastuullisuus on organisaatiolle luontaista, ei vastuita tarvitse välttämättä erikseen osoittaa. (Ladley 2012, s. 43-57.)

#### *Arviointi*

Tässä vaiheessa arvioidaan organisaation kykyjä datan hallinnointiin sekä kykyä olla hallinnoitu. Arviointi voidaan jakaa kolmeen kokonaisuuteen: kapasiteetti, kulttuuri ja yhteistyö. Kapasiteetilla tarkoitetaan organisaation kykyä muuttua, jota ei tule kuitenkaan sekoittaa haluun muuttua. Pelkkä halu muuttua ei riitä, vaan organisaatiolla on oltava valmiudet haluttuun muutokseen. Kulttuuri viittaa siihen, että jokainen organisaatio on yksilö. Yhteistyöllä tarkoitetaan organisaation kykyä luoda tiimejä, jotka koostuvat eri yksiköiden työntekijöistä. Tämä voidaan myös nähdä osana kulttuuria. (Ladley 2012, s. 43-57.)

### *Visio*

Tämän vaiheen tarkoituksena on osoittaa sidosryhmille ja johdolle datan hallinnoinnin merkitys sekä määritelmä. Tavoitteena on saavuttaa ymmärrys siitä, millainen datan hallinnoinnin ohjelma voisi olla omassa organisaatiossa. Hankkeen rajaus tulee kääntää datan hallinnoinnin määritelmäksi, joka sopii omalle organisaatiolle. (Ladley 2012, s. 43-57.)

### *Yhteensovittaminen ja liiketoiminta-arvo*

Tämän vaiheen tarkoituksena on linkittää datan hallinnointi suoraan liiketoiminnallisiin hyötyihin, kuten parempaan tuottavuuteen, laadukkaampaan päätöksentekoon tai parempiin analyyseihin (Ladley 2012, s. 43-57).

### *Toiminnallinen suunnittelu*

Tässä vaiheessa suunnitellaan datan hallinnoinnin käytännöt, linjaukset ja prosessit (Ladley 2012, s. 43-57).

### *Hallinnoinnin viitekehyksen suunnittelu*

Kun toiminnot on määritelty edellisessä vaiheessa, tulee ne siirtää organisaation viitekehykseen. Toisin sanoen organisaatiossa tulee tunnistaa datan omistajat ja isännöitsijät sekä muut olennaiset roolit ja vastuut. Roolit siis osoitetaan oikeille henkilöille. (Ladley 2012, s. 43-57.)

### *Road map*

Tässä vaiheessa suunnitellaan tapahtumat ja virstanpylväät, joilla organisaatio siirtyy hallinnoimattomasta tilasta hallinnoituun tilaan. Myös datan hallinnoinnin hanketta tukevat resurssit tulee määritellä. (Ladley 2012, s. 43-57.)

### *Roll-out ja ylläpito*

Datan hallinnoinnin ylläpito ei lopu ennen kuin se on täysin sisäistetty organisaatioon. Tässä vaiheessa kulttuurimuutoksen hallinta korostuu. Tämän vaiheen tavoitteena on myös varmistaa, että datan hallinnoinnin hanke saavuttaa sille asetetut tavoitteet. (Ladley 2012, s. 43-57.)

Ladleyn (2012) prosessimalli soveltuu hyvin datan hallinnoinnin perustamiseen ja käynnistämiseen organisaatiossa. Mallia voidaan myös soveltuvin osin käyttää datan hallinnoinnin jatkuvaan kehittämiseen. Se voi kuitenkin olla sellaisenaan liian raskas datan hallinnoinnin iteratiiviseen kehittämiseen, johon taas Karelin (2014b) malli (kuva 7) soveltuu mahdollisesti paremmin. Datan hallinnoinnin jatkuvassa kehittämisessä ei välttämättä

tarvitse perustella sidosryhmille tai johdolle sen merkitystä uudelleen tai sitoa datan hallinnointia uudestaan konkreettisiin liiketoimintahyötyihin. Jatkuvassa kehityksessä Ladleyn (2012) prosessi voidaan rajata esimerkiksi uuden kehitysiteraation rajaukseen, nykytilanteen arviointiin sekä käytäntöjen, linjausten, roolien ja road mapin päivittämiseen. Myös datan hallinnoinnin ylläpito on tärkeää, sillä täten voidaan varmistaa tavoitteiden saavuttaminen ja datan hallinnoinnin integroituminen organisaation jokapäiväiseen toimintaan. Huomattavaa on, että prosessin viimeisellä vaiheella ei ole selvää alku- ja loppumispäivää, vaan ylläpidon alettua se ei lopu ennen kuin datan hallinnointi on täysin sisäistetty (Ladley 2012, s. 56).

### 3.4 Isännöinti osana datan hallinnointia

Datan hallinnointiin liittyvässä kirjallisuudessa esiintyy usein termi isännöinti (*stewardship*). Isännöinti on toiminta jonkin resurssin hoitamiseksi, josta vastuu on osoitettu jollekin yksittäiselle henkilölle tai ryhmälle (Newman & Logan 2006, s. 3-4). Datan isännöintiä on siis datan hoitamisesta vastuussa olevan henkilön tai ryhmän toiminta sen laadun varmistamiseksi. Toisen näkemyksen mukaan datan isännöinti on datan hallinnoinnin lähestymistapa, joka vahvistaa vastuut tietoresurssien hallinnasta muiden puolesta sekä organisaation eduksi (McGilvray 2008, s. 53). Kolmannen määritelmän mukaan datan isännöinti koostuu ihmisistä, organisaatiosta ja prosesseista varmistaen, että osoitetut isännöitsijät ovat vastuussa hallitusta datasta (Plotkin 2014, s. 2). Lyhyesti sanottuna datan isännöitsijät ovat henkilöitä, jotka ovat vastuussa datan hoitamisesta ja hallitsemisesta (Sebastian-Coleman 2013, s. 20). Loshinin (2011, s. 122) listaamia datan isännöitsijän vastuuta ovat muun muassa:

- käyttäjäyhteisön tukeminen
- metadatan hallinta
- datan laatustandardien hallinta
- datan ylläpito
- datan laadun valvominen
- datan validointi
- tiedon jakaminen
- liiketoimintasääntöjen hallinta
- lähteiden hallinta
- pääsyn auktorisointi
- datan elinkaaren hallinta.

Datan isännöinnissä on toisin sanoen kyse datan hallinnoinnin operatiivisesta aspektista, jossa varsinainen jokapäiväinen datan hallinnointi tapahtuu. Ilman isännöintiä voi datan hallinnointi jäädä pelkäksi hyvien aikomusten viitekehyyksi, jota ei koskaan implementoida. (Plotkin 2014, s. 8.) On siis varmistettava, että datan hallinnointi pystytään jalkaut-

tamaan strategiselta tasolta yrityksen jokapäiväiseen toimintaan. Datan isännöinti ja isännöitsijät toimivat keinona siirtää korkean tason linjaukset käytäntöön. On kuitenkin vältettävä pelkkää henkilöiden nimittämistä johonkin titteliin, vaan keskityttävä vastuiden ja velvollisuuksien määrittelyyn (Ladley 2012, s. 126). On myös huomattava, että organisaatiossa voi olla useita datan isännöitsijöitä.

Edellä esiteltyjä näkemyksiä isännöinnistä ja isännöitsijöiden tehtävistä voidaan verrata Cohenin roolijakoon (kuva 4). Cohenin mallissa datan isännöinnin tehtävät on jaettu sekä isännöitsijälle että omistajalle. Cohenin roolijaossa on kaksi isännöinnin ulkopuolelle jäävää roolia, käyttäjä ja hallitsija, joilla on omat velvollisuudet isännöitsijöiden lisäksi. Organisaatiosta riippuen tehtävät voidaan joko jakaa isännöitsijöille ja omistajille tai pelkästään isännöitsijöille. Tällöin ei ole tarvetta käyttää omistaja-roolia. McGilvray (2008, s. 53-54) ei kannusta tiedon omistajuuteen, sillä tällöin ihmiset saattavat käyttäytyä kuin he ”omistaisivat” tiedon. Isännöinti-termin käyttö sen sijaan kannustaa vastuullisuuteen ja velvollisuuteen. Cheong & Changin malli (kuva 5) puolestaan sopii suoraan tässä alaluvussa esiteltyyn isännöitsijän rooliin. Cheong & Changin mallissa datan isännöitsijöitä voi olla useita, ja he toimivat datavastaavan alaisuudessa.

### 3.5 Yhteenveto

Datan hallinnointi on roolien, vastuiden, käytäntöjen ja prosessien määrittelyä tehokkaan ja proaktiivisen datan hallinnan varmistamiseksi. Sen avulla yritykset voivat kehittää datan hallintaansa toimivammaksi esimerkiksi kehittämällä käytäntöjä johdonmukaisemmiksi ja selventämällä rooleja ja vastuuta. Parhaimmillaan datan hallinnoinnilla voidaan myös saavuttaa parantunut datan laatu.

Datan hallinnoinnin systemaattisuutta eri yrityksissä voidaan kuvata datan hallinnoinnin erilaisilla kypsyysmalleilla. Mallit koostuvat kypsyystasoista. Nämä tasot taas sisältävät vaatimuksia, joiden tulee täytyä, jotta yritys pääsee kyseiselle tasolle. Kypsyysmallien avulla yritykset voivat selvittää oman datan hallinnointinsa systemaattisuuden ja määrittellä itselleen nykytason. Lisäksi kypsyysmallit mahdollistavat tavoitetasojen määrittelyn sekä yksityiskohtaisten tavoitteiden asettamisen. Kypsyysmalleja voidaan hyödyntää esimerkiksi datan hallinnan strategiaa luotaessa kuvaamaan nykytasoa sekä tavoitetilaa.

Datan hallinnoinnin organisoimisessa voidaan hyödyntää kirjallisuuden tarjoamia roolijakoja. Yleisiä rooleja ovat muun muassa datan isännöitsijä, käyttäjä, omistaja ja hallitsija. Roolien avulla voidaan selventää muun muassa datan hallintaan liittyvää vastuunjakoa. Tämä puolestaan edesauttaa tehokasta datan hallintaa. Selvät roolit myös tukevat IT- ja liiketoimintayksiköiden yhteistoimintaa. Roolit voidaan myös jakaa strategiaan, taktiikkiin ja operatiivisiin rooleihin. Strategisen tason roolit vastaavat organisaatioyksiköiden välisten ongelmien ratkomisesta sekä IT:n ja liiketoiminnan yhteensovittamisesta. Taktiset roolit muuttavat strategiset linjaukset taktiseksi suunnitelmaksi, auttavat operatiivista tasoa linjausten käyttöönotossa ja varmistavat, että data vastaa tarpeita. Datan käyttäjät ja



tuottajat toimivat operatiivisella tasolla. Heidän vastuulla on ongelmien esiin tuominen sekä liiketoimintavaatimusten tarkentaminen.

Datan hallinnoinnin viitekehykset ja prosessit auttavat yrityksiä ymmärtämään, mistä datan hallinnointi koostuu sekä huomioimaan kaikki datan hallinnoinnille olennaiset osa-alueet. Viitekehyksiä voidaan hyödyntää esimerkiksi datan hallinnan strategian luomisessa kypsyysmallien ohella. Niiden avulla voidaan datan hallinnan suunnitelmassa ottaa kantaa kaikkiin tarvittaviin näkökulmiin. Tämä puolestaan edesauttaa hankkeessa onnistumista ja tekee datan hallinnasta tehokkaampaa. Datan hallinnoinnin prosessimallit auttavat taas kuvaamaan hallinnoinnin eri vaiheita sekä korostavat sitä, että datan hallinnoinnissa ei ole kyse vain kertaluonteisesta hankkeesta. Niiden avulla esimerkiksi datan hallinnan kehittämisen suunnitelmassa voidaan huomioida kaikki olennaiset hankkeen vaiheet. Datan hallinnoinnin perustamisen prosessi taas auttaa yrityksiä yksityiskohtaisesti luomaan suunnitelman datan hallinnoinnin ohjelmalle. Tästä on erityisesti apua yrityksille, joissa ei ole vielä implementoitu datan hallinnointia.

Datan isännöinnin tarkoitus on varmistaa, että datan hallinnoinnin strategiset linjaukset jalkautetaan yrityksen jokapäiväisessä toiminnassa. Se on siis olennainen osa datan hallinnointia, ja yritysten tulee kiinnittää siihen huomiota, jotta siirtyminen hallinnoituun tilaan on mahdollisimman sujuvaa. Datan isännöinti voi myös auttaa vastaamaan muuttuviin liiketoimintatarpeisiin nopeammin ja joustavammin. Myös isännöinnissä on oleellista vastuiden ja velvollisuuksien selvä määrittely.

## 4. DATAN HALLINNAN JA LAADUN KEHITTÄMINEN

Datan laadun arvioinnin jälkeen siirrytään datan laadun kehittämiseen, jossa otetaan kantaa arvioinnissa esille nousseisiin datan laadun ongelmiin. Kuten luvussa 2.2 todettiin, datan laadun kehittämistä voidaan lähestyä data- tai prosessilähtöisesti. Tässä luvussa kuitenkin käsitellään vain prosessilähtöistä strategiaa. Luvun teoriaa hyödynnetään erityisesti haastattelutulosten analysoinnissa sekä kehitysehdotusten luomisessa.

Prosessilähtöinen datan laadun kehittäminen koostuu muun muassa erilaisista käytännöistä, kriittisistä onnistumistekijöistä sekä datan hallintaa ohjaavista prosesseista ja viitekehyksistä. Näiden tavoitteena on ohjata datan hallinnan kehittämishanketta sekä tarjota työkaluja ja suosituksia datan laadun ja hallinnan kehittämiseen. Viitekehykset mahdollistavat kaikkien hankkeelle tärkeiden osien ja sidosryhmien huomioimisen. Prosessimalit auttavat puolestaan tekemään datan hallinnan kehittämisestä järjestelmällisen ja jatkuvan prosessin. Kriittisten onnistumistekijöiden tunnistaminen mahdollistaa hankkeelle kriittisten asioiden huomioimisen. Erilaiset parhaiden käytäntöjen suositukset auttavat kehittämään datan hallinnan prosesseja tehokkaammiksi ja kestävämmiksi sekä selkeyttämään rooleja ja vastuita organisaatiossa.

### 4.1 Viitekehys datan hallinnan ongelmien löytämiseen

Datan hallinnan ongelmia voidaan selvittää ja niihin voidaan porautua hyödyntämällä erilaisia malleja ja viitekehyksiä, jotka ohjaavat ja standardoivat selvitysprosessia. Samoin kuin datan laadun arvioinnissa voidaan hyödyntää jotain vakiintunutta menetelmää, voidaan datan hallinnan ongelmia pyrkiä kartoittamaan jonkin viitekehyksen tai mallin avulla. Taulukossa 9 on esitetty McGilvrayn (2008, s. 21) POSMAD-viitekehys (*Plan, Obtain, Store and share, Maintain, Apply, Dispose*), jonka avulla voidaan löytää datan hallintaan liittyviä ongelmia ja haasteita. POSMAD-viitekehys kysymyksineen on esitetty tarkemmin liitteessä C.

**Taulukko 9.** POSMAD-viitekehyksen matriisi (McGilvray 2008, s. 21)

	Suunnittelu	Kerääminen	Tallettaminen ja jakaminen	Ylläpito	Käyttö	Hävittäminen
Data (mitä)						
Prosessit (miten)						
Ihmiset ja organisaatiot (kuka)						
Teknologia (miten)						

POSMAD-viitekehyksessä datan hallintaa tarkastellaan datan elinkaaren kautta. Datan elinkaari koostuu vaiheista suunnittelu/määrittely, kerääminen, tallettaminen ja jakaminen, ylläpito, käyttö sekä hävittäminen. Jokaista elinkaaren vaihetta tarkastellaan neljällä eri tasolla, jotka pyrkivät vastaamaan kysymyksiin mitä, miten ja kuka. Näin tarkastelussa voidaan kattaa mahdollisimman monta datan hallinnan osa-aluetta. McGilvrayn (2008, s. 19) mukaan mitä tahansa resurssia hallittaessa on tärkeää ymmärtää elinkaaren konsepti; jotta resursseista voidaan saada kaikki hyöty irti, tulee niitä hallita niiden koko elinkaaren läpi.

POSMAD-viitekehyksessä datan hallintaa tarkastellaan vuorovaikutusmatriisin avulla, jossa komponentit data, prosessit, ihmiset ja organisaatiot sekä teknologia ovat vuorovaikutuksessa elinkaaren eri vaiheiden kanssa. Jokaiseen matriisiin soluun liittyy kyseistä vuorovaikutusta koskevia kysymyksiä, joiden avulla voidaan paremmin ymmärtää tämänhetkistä tilannetta sekä suunnittelemaan uusia prosesseja. (McGilvray 2008, s. 20.) Matriisin solujen mallikysymykset on esitetty liitteen C tarkemmassa POSMAD-viitekehyksen esityksessä. Mallikysymyksiä voidaan täydentää lisäkysymyksillä, jotka soveltuvat juuri kohdeyrityksen kontekstiin ja olosuhteisiin. Näin kehystä voidaan kohdentaa kohdeyritykselle sopivammaksi.

## 4.2 Hallinnoinnin ja laadun kehittämisen kriittiset onnistumistekijät

Kun organisaatio lähtee toteuttamaan jotain suurta tai tärkeää hanketta, voi se käyttää kriittisiä onnistumistekijöitä apuna hankkeelle olennaisten ja kriittisten asioiden huomioimisessa. Erilaisille hankkeille on omat tyypilliset onnistumistekijät, joihin tulee keskittyä ja jotka tulee tiedostaa hankkeen edetessä. Kriittiset onnistumistekijät (*critical success factors*) ovat erityispiirteitä, ehtoja tai muuttujia, joilla voi oikein ylläpidettyinä tai hallituina olla merkittävä vaikutus jollain tietyllä toimialalla kilpailevan organisaation menestymiseen (Leidecker & Bruno 1984; Lucas 2010 mukaan). Taulukkoon 10 on koottu kirjallisuudessa esiteltyjä datan hallinnoinnin ja datan laadun kehittämishankkeiden kriittisiä onnistumistekijöitä. Esiteltyt kriittiset onnistumistekijät on järjestetty sen mukaan, kuinka monessa teoksessa kyseiseen onnistumistekijään on viitattu.

**Taulukko 10.** *Datan hallinnoinnin ja laadun kehittämisen kriittisiä onnistumistekijöitä*

Kriittinen onnistumistekijä	Xu 2003	Marinos 2004	Adler 2007	Tee et al. 2007	Lucas 2010	Panian 2010	Ladley 2012
Johdon sitoutuminen/tuki	X	X	X	X		X	
Koulutus, tietoisuus ja ymmärrys	X	X		X			
Kommunikointi/motivointi	X			X	X		
Seuranta/monitorointi		X	X			X	
Strategia/visio	X		X				
Selkeä roolijako		X				X	
Organisaatiokulttuuri ja sen muutoksen hallinta	X						X
Datan laadun/hallinnan puolestapuhuja ( <i>champion</i> )				X			
Henkilöstön kyvykkyys Keskijohdon sitoutuminen/tuki Input-tarkistukset	X						
Standardit Mittarit		X					
Liiketoiminnan ja IT:n yhteistoiminta						X	

Tärkein onnistumistekijä datan laadun kehittämishankkeissa on johdon tuki ja sitoutuminen projektiin. Johdon tuki toimii esimerkkinä muulle organisaatiolle ja viestii hankkeen liiketoiminnallisesta tärkeydestä. Ylimmän johdon sitoutuminen voi toimia mittarina koko organisaation sitoutumiselle datan laadun hallintaan (Xu 2003, s. 185). Datan hallinnointi vaatii, että johto tunnistaa datan hallinnoinnin arvon, jotta hanke voidaan sitoa konkreettisiin liiketoimintatavoitteisiin (Panian 2010, s. 944). Lisäksi hankkeeseen sitoutunut johto voi auttaa luomaan yhteisymmärrystä eri sidosryhmien välillä (Adler 2007), jalkauttamaan strategisia tavoitteita käytännön tekemiseksi sekä välttämään poliittista kädenvääntöä (Niemi & Kontra 2012, s. 4). Johdon sitouttaminen voi olla kuitenkin haastavaa, sillä johto on yleensä kiinnostunut enemmän liiketoiminnan mittareista. Datan laadun arvo realisoituu vasta negatiivisesti ylimääraisten kustannusten muodossa. Tämän vuoksi johdolle on pystyttävä osoittamaan hyvälaatuisen datan tuottama arvo sekä korostaa datan laadun merkitystä osana liiketoiminnan laatua. (Laatikainen & Niemi 2012, s. 7-8.)

Toinen tärkeä onnistumistekijä on koulutus, tietoisuus ja ymmärrys. Datan hallinnoinnin ja kehityshankkeen parissa työskentelevien henkilöiden tulee ymmärtää datan laatua ja sen merkitys. Yhteiset sovitut käytännöt datan hallinnassa tulee kouluttaa koko henkilöstölle, jotta käytännöt voidaan yhtenäistää kautta organisaation. Koulutusohjelma voi auttaa voimaannuttamaan henkilöstöä, mikä puolestaan johtaa parempaan laadun hallintaan (Xu 2003, s. 185). Henkilöstön kouluttaminen on yksi tärkeimmistä strategioista ylläpitämään datan hallinnan kestävyyttä (Marinos 2004).

Tietoisuutta datan hallinnoinnista ja laadun kehittämishankkeesta tulee aktiivisesti jakaa ja levittää koko organisaatiolle. Tämä taas linkittyy hankkeesta kommunikointiin, joka on myös yksi kriittinen onnistumistekijä. Yksi merkittävä tekijä datan laadun hallinnassa on työntekijöiden motivointi innovatiivisilla kommunikointimenetelmillä (Lucas 2010, s. 1).

Myös hankkeen seurannalla ja monitoroinnilla on suuri merkitys. Datan hallinnoinnin hankkeen arvo ei toteudu, jos organisaatiossa ei varmisteta, että sovittuja käytäntöjä todella noudatetaan (Marinos 2004). Organisaatio voi varmistaa tehokkaan datan hallinnoinnin ottamalla käyttöön johdonmukaisen menetelmän parhaiden käytäntöjen ja teknologioiden dokumentointiin (Adler 2007). Toisin sanoen, datan laadun kehittämishanke on jäänyt kesken, jos sen hyötyjä ei olla todennettu.

Organisaation tulisi luoda strategia tai visio datan hallinnoinnin tulevaisuudentilalle (Adler 2007). Tämä asettaa tavoitteet ja ohjaa hanketta. Olennainen vaatimus datan hallinnoinnin hankkeelle on roolien ja vastuiden tarkka määrittely (Panian 2010, s. 943). Näin organisaatiossa kannustetaan vastuullisuuteen. Datan hallinnoinnin hankkeessa tulee myös hallita organisaation muutosta (Ladley 2012, s. 32). Jos datan laatu on huonoa, on vanhoja toimintatapoja ja käytäntöjä muutettava. Lisäksi hankkeen puolestapuhuja (*champion*) on yksi onnistumistekijä (Tee et al. 2007, s. 351). Puolestapuhuja kannustaa organisaatiota ja johtoa sitoutumaan hankkeeseen. Organisaation liiketoimintapuolen ja IT-puolen yhteistoiminta on myös kriittistä hankkeen onnistumiselle (Panian 2010, s. 944). Liiketoimintapuoli ottaa vastuun datasta, ja IT-puoli auttaa teknologisten aspektien implementoinnissa. Myös Malange et al. (2015, s. 1) ja Page (2011, s. 18) korostavat liiketoiminnan ja IT:n yhteistyötä, mutta eivät kuitenkaan mainitse sitä kriittisenä onnistumistekijänä.

### 4.3 Organisatoriset käytännöt datan laadun parantamiseksi

Organisaation kehittyessä sen lähestyminen käytäntöjen noudattamiseen muuttuu epävirallisesta lähestymisestä sellaiseen, jossa liiketoiminta-aktiviteetit ja tietoon liittyvät käytännöt on integroitu (Loshin 2011, s. 46). Käytännöt ovat siis luonnollinen osa yrityksen toimintaa. Taulukossa 11 on kypsyyssmalli, joka esittää datan hallintaan liittyvien käytäntöjen kehitystä.

**Taulukko 11.** *Datan laadun hallinnan käytäntöjen kypsyysmalli (Loshin 2011, s. 47)*

Taso	Kuvaus
Varhainen	<ul style="list-style-type: none"> <li>Käytännöt ovat epävirallisia</li> <li>Käytäntöjä ei ole dokumentoitu</li> <li>Henkilöstö tekee toisteisia toimintoja ilman koordinoitua</li> </ul>
Toistettava	<ul style="list-style-type: none"> <li>Organisaatio pyrkii yhdistämään datan yhteen lähteeseen</li> <li>Yksityisyyteen ja käytön rajoitukseen liittyvät käytännöt ovat iskostettuja</li> <li>Varhaiset käytännöt dataan liittyvien ongelmien ratkomiseksi määritelly</li> </ul>
Määriteltä	<ul style="list-style-type: none"> <li>Räätälöidyt ohjeet hallinnan tavoitteiden luomiselle on kehitetty</li> <li>Sertifointiprosessi datalähteiden arvioinnille on otettu käyttöön</li> <li>Datan laadun harjoittajat kaappaavat parhaita käytäntöjä</li> <li>Datan laadun palvelutasosopimukset määritelly käytäntöjen seuraamiseksi</li> </ul>
Hallittu	<ul style="list-style-type: none"> <li>Käytännöt on perustettu ja niitä koordinoidaan kautta organisaation</li> <li>Datan laadun hallinta perustuu käytäntöihin</li> <li>Datan laadun käytännöt ohjaavat suorituskvyn hallintaa</li> <li>Datan laadun palvelutasosopimusta käytetään käytäntöjen seuraamiseksi</li> </ul>
Optimoitu	<ul style="list-style-type: none"> <li>Datan laadun käytäntöjen noudattamatta jättämisestä ilmoitetaan automaattisesti</li> <li>Itsehallinnoiva järjestelmä</li> </ul>

Loshinin (2011, s. 47) käytäntöjen kypsyysmallissa varhaisilla tasoilla korostetaan käytäntöjen epävirallisuutta ja kehittyneemmällä tasoilla korostetaan käytäntöjen luonnollisuutta ja datan hallinnan automatisointia. Vaikka malli ei tarjoa parhaita käytäntöjä yrityksen datan hallinnan kehittämiseen, voi se toimia hyvänä mallina arvioida yrityksen datan hallinnan käytäntöjen nykytilaa. Toisaalta malli voi toimia tavoitteena käytännöille datan hallinnan kehittämishankkeessa. Yritys voi asettaa jonkin mallin tason datan hallinnan käytäntöjen tavoitetilaksi. Datan hallinnan kehittämisen parhaita käytäntöjä eri teoksista on koottu taulukkoon 12. Parhaat käytännöt ovat menetelmä tai tekniikka, joka on jatkuvasti tuottanut parempia tuloksia kuin muilla menetelmillä saadut tulokset, jolloin sitä voidaan pitää benchmarkina (BusinessDictionary 2015a). Toisen määritelmän mukaan parhaat käytännöt ovat menetelmien kokoelma, joka on virallisesti hyväksytty olevan paras tapa toimia jollain tietyllä alalla tai tietyssä liiketoiminnassa (Cambridge Dictionaries Online 2015). Tunnistamalla datan laadun ja hallinnan kehityksen parhaita käytäntöjä, voidaan näihin kiinnittää huomiota, kun organisaatiossa aletaan suunnitella ja toteuttaa dataan liittyviä kehityshankkeita.

**Taulukko 12.** *Kokoelma datan laadun parhaita käytäntöjä datan hallinnan kehittämiseksi*

Käytäntö	Lähteet
Palkitsemisjärjestelmä vahvistamaan datan ja sen laadun merkitystä	Khatri & Brown 2010; Storey et al. 2012
Lähestymisen datan hallinnointiin tulee olla holistinen ja ottaa kantaa ihmisiin, prosesseihin ja teknologiaan. Datan hallinnoinnissa on kyse niin johtamisesta, kommunikoinnista kuin teknisestä integroinnista.	
Selkeät datan laadun tavoitteet, esimerkiksi tietojärjestelmästandardien kautta	Todd 2008; Friedman 2006
Datan käsittely jätetään päätöksentekijälle. Näin prosessoinnin tekee henkilö, joka varmasti arvostaa kyseisen prosessoitavan datan laatua. Dataa ei prosessoi henkilöt, joille laatu on vain lähes abstrakti asia.	
Työntekijöiden voimaannuttaminen: työntekijä mahdollisimman lähellä prosessin alkupäätä tekee päätökset.	
Tiimit: Korvataan hierarkkinen organisaatio käyttämällä tiimejä, mikä auttaa organisaation muutoksessa. Tiimit helpottavat vuorovaikutusta yksilöiden välillä. Tieto toisten laatu päätöksistä voi muuttaa yksilön tekemät päätökset lähemmäs koko organisaation linjausta.	
Suorituskykyä mitataan selkeillä ja suorilla ehdoilla, työntekijöille annetaan palautetta, mittarit liittyvät suoraan datan laatuun ja tärkeisiin laadun ulottuvuuksiin	
Datan omistajuuden käytännöt	
Liiketoimintajohtajien komitea kehittämään ja valvomaan dataan liittyviä käytäntöjä	Storey et al. 2012
Komitea, joka ohjaa liiketoiminnan ja IT:n yhteistyötä	
Tiedoksiannot ja intranet-utiset toimivat käytäntöihin, prosesseihin ja tavoitteisiin liittyvien päätösten kommunikointikeinona	
Business case ja mittarit on määritelty aikaisessa vaiheessa.	Khatri & Brown 2010
Datan hallinnoinnin implementointi on matka, jossa on tietyt vaiheet, jotka vuorostaan kuvaavat organisaation kehitystä datan hallinnoinnissa.	
Datan hallinnointiin liittyville hyödyille, aikajanoille ja kyvykkyyksille on asetettu realistiset odotukset. Esimerkiksi datan hallinnointi asettaa standardit datan laadulle, mutta ei tee itse datan korjausta.	
Datan hallinnointiin ryhdytään kattavan datan hallinnan näkökulmasta joka ottaa kantaa myös data-arkkitehtuuriin, metadataan, MDM:ään, datan laatuun ja turvallisuuteen.	

*Datan hallinnointia tulee lähestyä holistisesti*, mikä tarkoittaa niin ihmisten, prosessien kuin teknologioiden huomioimista (Friedman 2006, s. 4; Todd 2008, s. 31). Datan hallinnoinnissa on tämän vuoksi kyse johtamisesta, hyvästä hallitsemisesta sekä teknisestä integroinnista (Todd 2008, s. 31).

Khatri & Brown (2010, s. 152) sekä Storey et al. (2012, s. 440) korostavat toimivaa datan laadun *palkitsemisjärjestelmää*. Sen tarkoituksena on vahvistaa datan ja datan laadun merkitystä voimavarana organisaatiossa. Sen avulla työntekijöitä voidaan kannustaa datan tehokkaampaan hallintaan ja parempaan laatuun.

Palkitsemisjärjestelmä linkittyy myös *selkeisiin laadun tavoitteisiin*. Tavoitteet voidaan asettaa esimerkiksi laatua koskevien tietojärjestelmästandardien avulla (Storey et al. 2012, s. 439). Palkitsemisjärjestelmä taas voi toimia sen perusteella, kuinka hyvin näihin tavoitteisiin päästään.

*Datan omistajuuden käytännöt* sisältää muun muassa IT-resurssien hajauttamisen. Toisin sanoen, osasto, joka käyttää eniten jotain tiettyä dataa, myös hallitsee ja omistaa datan. Tällöin se voi itse toteuttaa omia laatuvaatimuksiaan. (Storey et al. 2012, s. 440.) IT-resurssien hajauttamiseen tulee kuitenkin suhtautua varautuneesti. Jos datan hallintaa hajautetaan liikaa eri osastoille, voi organisaation data päätyä useisiin tietojärjestelmiin tai tietokantoihin, mikä taas vaikeuttaa datan hallintaa. Myös käytäntöjen yhtenäistäminen kautta koko organisaation vaikeutuu.

*Työntekijöiden voimaannuttamisessa* on kyse siitä, että päätäntävaltaa siirretään mahdollisimman lähelle prosessin alkupäätä. Työntekijät, jotka keräävät ja ymmärtävät dataa, tekevät enemmän päätöksentekoon liittyvää prosessointia. Työntekijöiden voimaannuttaminen vähentää datan siirtoa henkilöltä toiselle, mikä taas lisää datan laatua. (Storey et al. 2012, s. 438-439.) Khatri & Brown (2010, s. 152) puolestaan ehdottavat erilaisten *komiteoiden* käyttöönottoa ja korostavat *kommunikoinnin merkitystä*. Komiteat voivat esimerkiksi hyväksyä IT-projektiehdotuksia sekä kehittää ja valvoa dataan liittyviä käytäntöjä ja niiden noudattamista sekä auttaa liiketoiminta- ja IT-yksiköiden yhteistyössä. Tiedotukset yrityksen sisäisten nettisivujen kautta voivat toimia keinona jakaa tietoa käytännöistä, prosesseista sekä datan hallinnoinnin tavoitteista. (Khatri & Brown 2010, s. 152.)

Todd (2008, s. 31) listaa ominaisuuksia, jotka ovat olleet yhteisiä datan hallinnoinnissa kehittyneillä yrityksillä. *Business case* ohjaa datan hallinnoinnin implementoinnin prioriteetteja ja mittarit mahdollistavat liiketoiminnallisten hyötyjen todentamisen. Hän mainitsee myös tärkeiksi ominaisuuksiksi *realistiset odotukset* sekä *datan hallinnoinnin liittäminen muihin hallitsemisen lähestymistapoihin*, kuten datan laatuun ja MDM:ään (*Master Data Management*).

#### 4.4 Datan laadun kehittämisen prosessi

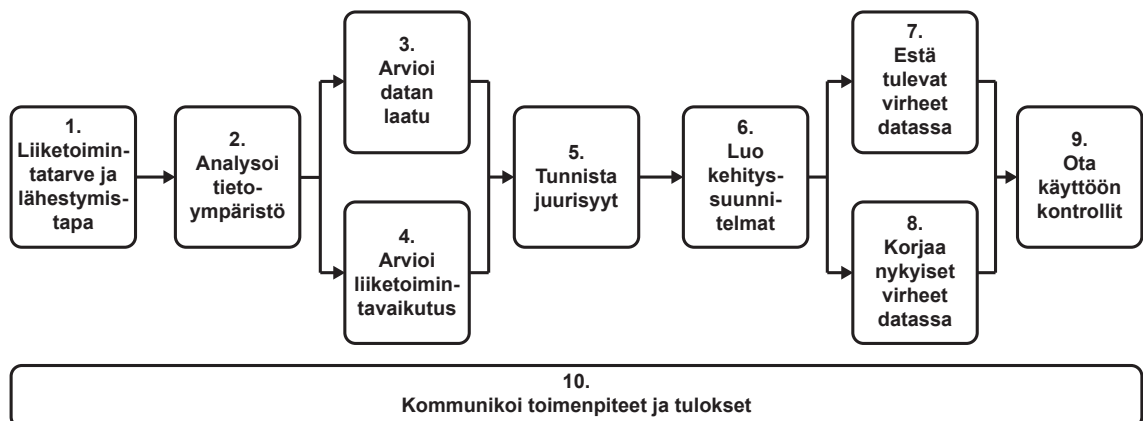
Parhaimpien käytäntöjen lisäksi datan hallinnan ja laadun kehittämistä voidaan ajatella prosessimaisesti tai erilaisten pykälämallien avulla. Batini et al. (2009, s. 4) listaavat datan laadun kehittämisen askeleet:

- *kustannusten arviointi*
  - datan laadun suorat ja epäsuorat kustannukset
- *prosessivastuiden jakaminen*
  - prosessien omistajien tunnistaminen ja heidän vastuiden määrittelemineen datan tuotannossa ja hallitsemisessa



- *datavastuiden jakaminen*
  - datan omistajien tunnistaminen ja heidän vastuiden määritteleminen
- *virhesyiden tunnistaminen*
  - laatuongelmien syiden tunnistaminen
- *strategioiden ja tekniikoiden valinta*
  - laatutavoitteisiin ja budjettirajoitteisiin sopivien datan kehityksen strategioiden ja tekniikoiden tunnistaminen
- *datan kehityksen ratkaisuiden suunnittelu*
  - tehokkaimman strategian ja siihen liittyvien tekniikoiden ja työkalujen valinta datan laadun parantamiseksi
- *prosessienhallinta*
  - datan tuotantoprosessin tarkastuspisteiden määritteleminen datan laadun seuraamiseksi
- *prosessien uudelleen suunnittelu*
  - prosessien kehitystoimenpiteiden määritteleminen
- *kehityksen hallinta*
  - uudet organisatoriset säännöt datan laadulle
- *kehityksen seuranta*
  - määritellään säännölliset seurantatoimenpiteet, jotka tarjoavat palautetta kehitystoimenpiteistä ja mahdollistavat niiden hienosäädön.

Datan laadun kehittämisen askeleilla on yhteistä datan hallinnoinnin prosessin kanssa se, että molemmissa on tärkeää selkeiden vastuiden jakaminen ja kehityksen seuranta. Datan laadun kehittäminen voidaan myös esittää suoraan prosessimaisesti. Kuvassa 9 on McGilvrays (2008, s. 64) prosessi datan laadun kehittämiseen.



**Kuva 9.** Kymmenen askeleen prosessi (mukaillen McGilvrays 2008, s. 64)

McGilvrayn (2008, s. 292) prosessimalli datan laadun arviointiin ja kehittämiseen koostuu kymmenestä vaiheesta:

1. *Liiketoimintatarve ja lähestymistapa*
  - Määrittele ongelma, mahdollisuus tai tavoite ohjaamaan kaikkea projektin työtä. Palaa tähän askeleeseen pitääksesi maalin koko ajan kirkkaana mielessä.
2. *Analysoi tietoympäristö*
  - Kerää, koosta ja analysoi tietoa tämänhetkisestä tilanteesta ja tietoympäristöstä. Dokumentoi ja verifioi tiedon elinkaari. Tämä toimii pohjana tuleville askelille, auttaa löytämään juurisyitä sekä varmistaa, että olennaista dataa arvioidaan. Luo datan kaappaamisen ja arvioinnin suunnitelma.
3. *Arvioi datan laatu*
  - Arvioi datan laatua sopivien laatu-ulottuvuuksien kautta. Arvioinnin tulokset toimivat pohjana tuleville askelille, kuten juurisyiden tunnistaminen sekä parannukset ja datan korjaukset.
4. *Arvioi liiketoimintavaikutus*
  - Arvioi huonolaatuisen datan vaikutusta liiketoimintaan käyttäen eri tekniikoita. Tämä askel tuottaa business casen kehitykselle, kannustaa tukemaan datan laadun parantamista sekä auttaa valitsemaan sopivat investoinnit.
5. *Tunnista juurisyöt*
  - Tunnista ja priorisoi datan laadun todelliset syyt ja kehitä spesifiset suositukset niiden korjaamiseksi.
6. *Luo kehityssuunnitelmat*
  - Viimeistele spesifiset suositukset toimenpiteitä varten. Kehitä ja suorita kehityssuunnitelmat suositusten perusteella.
7. *Estä tulevat virheet datassa*
  - Ota käyttöön ratkaisut, jotka ottavat kantaa datan laadun juurisyihin.
8. *Korjaa nykyiset virheet datassa*
  - Ota käyttöön sopivat datan korjauksen toimenpiteet.
9. *Ota käyttöön kontrollit*
  - Seuraa ja verifioi, että kehitystoimenpiteet on todella implementoitu. Ylläpidä kehitettyjä tuloksia standardisoimalla, dokumentoimalla ja jatkuvasti seuraamalla onnistuneita käyttöönottoja.
10. *Kommunikoi toimenpiteet ja tulokset*
  - Dokumentoi ja kommunikoi laatu-testien tulokset, tehdyt parannukset sekä näiden parannusten tulokset. Kommunikointi on niin tärkeää, että se on osa jokaista prosessin askelta.

Kymmenen askeleen prosessissa voidaan lisäksi hyödyntää POSMAD-viitekehystä (ks. luku 4.1). Esimerkiksi laadun arvioinnissa voidaan tarkastella kaikkia datan elinkaaren

vaiheita, sillä datan laatu voi muuttua kaikissa vaiheissa. Liiketoimintavaikutusten arvioinnissa ja business casen tuottamisessa voidaan taas keskittyä käyttö-vaiheeseen. Dataa käytetään aina täyttämään jokin liiketoimintatavoite. (McGilvray 2008, s. 93.) POSMAD-viitekehyksen huomioiminen datan laadun kehittämisen hankkeessa korostaa datan elinkaariajattelua sekä organisaation eri tasoja (data, prosessit, ihmiset, teknologia). Tämä puolestaan mahdollistaa kokonaisvaltaisemman ajattelun, jossa kaikki näkökulmat ja vaikutukset datan laatuun huomioidaan.

Ensimmäisen ja viimeisen vaiheen merkitystä tulee korostaa. Datan laadun kehittämishanke tulee perustella ja sille tulee asettaa selkeät tavoitteet. Hankkeen tulee lähteä liiketoimintatarpeesta ja sen tarkoituksena tulee olla tuottaa yritykselle liiketoiminnallista hyötyä. Selkeät tavoitteet myös motivoivat henkilöstöä; organisaatiolla tulee olla mielessä jokin selkeä tavoitetila, johon työntekijät yhdessä pyrkivät. Parantunut datan laatu ei vielä riitä tavoitteen määrittelyksi, vaan määrittelyssä tulee täsmentää organisaatiolle kriittisiä laatu-ulottuvuuksia sekä laadun mittareita. Näiden avulla parannukset datan laadussa pystytään todentamaan ja mittamaan.

Toisaalta myös hankkeesta kommunikoinnilla on suuri merkitys. Ilman tehokasta kommunikointia työntekijät voivat etäännyä hankkeesta, mikä taas näkyy negatiivisesti motivaatiossa. Tiedottamisesta tulee pitää siis huolta koko hankkeen ajan. Kommunikointi on tärkeää myös käytäntöjen yhtenäistämisessä ja tuloksista tiedottamisessa. Muutoksista datan käsittelyyn liittyvistä käytännöistä ja prosesseista tulee tiedottaa koko organisaatiolle, jotta ne voidaan yhtenäistää kautta linjan. Tuloksista tulee tiedottaa myös koko yritykselle, jotta kaikki tietävät, missä on onnistuttu ja missä on kehitettävää.

## 4.5 Yhteenveto

Datan hallinnan kehittäminen lähtee hankkeen suunnittelusta. Suunnitteluun liittyviä näkökulmia on esitetty enemmän luvussa 3. Suunnittelua seuraa kuitenkin olemassa olevien ongelmien ja haasteiden tunnistaminen ja kartoitus. Tässä yritykset voivat hyödyntää esimerkiksi POSMAD-viitekehyksen kaltaisia apuvälineitä. POSMAD-viitekehys mahdollistaa datan hallinnan ongelmien sijoittamisen johonkin vaiheeseen data elinkaarta sekä jollekin tietylle organisaatiotasolle. Kun ongelmat voidaan sijoittaa johonkin tiettyyn vaiheeseen tai jollekin tietylle organisaatiotasolle, voi niiden korjaaminen olla huomattavasti helpompaa. Ongelmien sijoittamisella POSMAD-matriisiin, voidaan myös tarkastella, painottuvatko ongelmat jollekin tasolle tai johonkin vaiheeseen. Viitekehys siis helpottaa ongelmien tunnistamista, mutta se auttaa myös ongelmien analysoinnissa.

Kuitenkin jo hankkeen suunnitteluvaiheessa yritykset voivat hyödyntää kriittisiä onnistumistekijöitä tunnistessaan omalle hankkeelle kriittiset näkökulmat. Ne auttavat huomioimaan hankkeen kannalta olennaiset piirteet, joilla on merkittävä vaikutus hankkeessa onnistumiseen. Datan hallinnan ja laadun kehittämisen hankkeen selvästi kriittisin onnistumistekijä on johdon tuki ja sitoutuminen. Kriittiset onnistumistekijät eivät sinänsä ole

suoraan datan hallinnan tai laadun kehittämisen keinoja, mutta ne linkittyvät oleellisesti datan hallinnan ja laadun kehittämisen hankkeen menestykseen. Tämän vuoksi ne voidaan nähdä tärkeänä työkaluna datan hallinnan kehittämisessä.

Datan hallinnan ja laadun parhaat käytännöt tarjoavat yrityksille onnistumistekijöitä konkreettisempia ehdotuksia toiminnan kehittämiseen. Parhaat käytännöt ovat yleisesti hyväksyttyjä parhaita toimintatapoja tietyssä toiminnassa tai tietyllä alalla. Kun esimerkiksi datan hallinnan organisoimista, roolien vastuita sekä uusia käytäntöjä ja prosesseja suunnitellaan, yritykset voivat tukeutua parhaisiin käytäntöihin. Yleisimmin mainitut parhaat käytännöt ovat datan ja sen laadun merkitystä vahvistava palkitsemisjärjestelmä sekä holistinen lähestyminen datan hallinnointiin.

Datan laadun kehittämisen prosessit toimivat ohjenuorana laadun kehittämisen hankkeissa. Ne jakavat hankkeen selkeisiin osakokonaisuuksiin ja vaiheisiin sekä varmistavat, että nämä kaikki vaiheet otetaan huomioon hankkeen edetessä. Hanketta suunniteltaessa voidaan näitä prosessimalleja käyttää apuna määriteltäessä hankkeen eri vaiheet. Yksinkertaistetusti prosessi etenee seuraavasti. Prosessit lähtevät liiketoimintanäkökulmasta; laadun kehittämiseksi pitää olla liiketoiminnallinen tarve tai ajuri. Tämän jälkeen arvioidaan datan laadun nykytilanne ja sen vaikutus sekä korjataan ongelmia tuottavat prosessit. Viimeiseksi korjataan olemassa olevat ongelmat datassa ja otetaan käyttöön seurantakeinot. On tärkeää huomata, että nykyisten ongelmien korjaaminen tapahtuu vasta ongelmia tuottavien prosessien korjaamisen jälkeen. Ensin varmistetaan, että dataa tuottavat prosessit tuottavat korkealaatuista dataa. Vasta tämän jälkeen aletaan korjata olemassa olevan datan ongelmia. Näin vältetään tilanteelta, jossa dataa joudutaan korjaamaan sitä mukaa kuin uutta dataa tuotetaan.

## 5. MENETELMÄT

Tässä luvussa käsitellään työssä käytettyjen empiiristen menetelmien toteutus. Työssä on kaksi arviointikokonaisuutta: datan laadun arviointi sekä datan hallinnan arviointi. Molemmissa kokonaisuuksissa käytetään haastattelua empiirisenä tiedonkeruumenetelmänä. Datan laadun arvioinnissa käytetään soveltuvin osin AIMQ-menetelmää. Datan hallinnan arvioinnissa taas hyödynnetään POSMAD-viitekehystä tulosten analysoinnissa. Molempia arviointikokonaisuuksia tarkastellaan tässä luvussa niiden vaiheiden mukaan.

Datan laadun arvioinnin lopputuloksena saadaan kokonaiskuva tarkastellun datan laadusta. Näin tuloksista voidaan löytää datan laadun ulottuvuudet, jotka eivät ole halutulla tasolla tai joissa on huomattavasti ongelmia. Datan hallinnan arvioinnin lopputulos on kuvaus datan hallinnan ongelmista ja haasteista kohdeyrityksessä. Lisäksi datan hallinnan arvioinnissa määritellään datan hallinnan nyky- ja tavoitetilanne. Nämä tulokset toimivat pohjana kehitysehdotuksille, jotka ovat tämän työn pääasiallinen tulos kohdeyrityksen kannalta.

### 5.1 Datan laadun arviointi

Tässä työssä tehtävä datan laadun arviointimenetelmä voidaan kuvata luvussa 2.2 esitellyjen vertailukohtien avulla. Käytetystä menetelmästä voidaan tarkastella sen vaiheita, strategiaa, mittareita sekä tarkasteltuja datatyppejä ja tietojärjestelmätyyppejä. Taulukossa 13 on esitetty menetelmän vaiheet sekä niiden kuvaukset.

Data-analyysivaihe on toteutettu pääasiallisesti jo ennen varsinaista tutkimusta. Tänä aikana tämän työn tekijällä on ollut noin kymmenen kuukautta aikaa tutustua kohdeyrityksen dataan ja sen hallintaan liittyviin käytäntöihin. Kriittiset osa-alueet on myös tunnistettu kohdeyrityksen kanssa jo tutkimuksen suunnitteluvaiheessa. Tarkasteltava data on kemian ja hydrologian tutkimustulokset sekä tutkimustietokannassa että sen ulkopuolella.

Työssä käytetään AIMQ-menetelmän kyselylomaketta soveltuvin osin. AIMQ-menetelmän kyselylomake on esitetty yksityiskohtaisesti liitteessä B. Datan laadun arviointiin liittyvä aineisto kerätään haastattelujen avulla. Haastatteluihin valitaan kohdeyrityksen työntekijöitä, jotka omassa työnkuvassaan työskentelevät eri datan hallinnan rooleissa sekä tutkimusdatan elinkaaren eri vaiheissa. AIMQ-kyselylomakkeen väitteisiin ei kuitenkaan kerätä numeerisia arvosanoja vaan vapaamuotoista sanallista palautetta. Näin voidaan saada tarkempaa tietoa siitä, mitä ongelmia kuhunkin datan laadun ulottuvuuteen liittyy. AIMQ-menetelmän väitteet on myös muunneltu kysymyksiksi, jotka on esitetty haastattelupohjassa liitteessä A.

Haastattelut ovat puolistrukturoituja. Haastateltavilta kysytään ennalta määritetyt kysymykset, jotka on esitetty liitteessä A. Näiden kysymysten lisäksi haastateltaville voidaan esittää tarpeen mukaan tarkentavia tai laajentavia kysymyksiä. Näin haastateltavilta voidaan saada tärkeitä näkemyksiä tai tarkennuksia, joita ei olisi välttämättä saatu strukturoidussa haastattelussa.

**Taulukko 13.** *Datan laadun arvioinnin vaiheet*

Vaihe	Kuvaus
Data-analyysi	Tutkija on tutustunut dataan ja sen hallintaan liittyviin käytäntöihin työskentelemällä datan hallinnan parissa kohdeyrityksessä noin kymmenen kuukauden ajan.
Kriittisten alueiden tunnistaminen	Kohdeyrityksen kanssa on tunnistettu kriittiset datapaketit ja rajattu tarkastelu niihin. Tarkasteltava data on rajattu kemian ja hydrologian tutkimusdataan, joka sijaitsee tutkimusdatatietokannassa ja sen ulkopuolella.
Valitaan relevantit ulottuvuudet sekä määritellään kysymykset, mittarit ja tekniikat	<p>Työssä käytetään AIMQ-menetelmän kyselylomaketta soveltuvin osin. Tarkasteltavat ulottuvuudet ovat täten:</p> <ul style="list-style-type: none"> <li>• saatavuus</li> <li>• sopiva määrä</li> <li>• luotettavuus</li> <li>• täydellisyys</li> <li>• tiivis esitystapa</li> <li>• johdonmukainen esitystapa</li> <li>• helppokäyttöisyys</li> <li>• oikeellisuus/tarkkuus</li> <li>• tulkittavuus</li> <li>• objektiivisuus</li> <li>• relevanssi/hyödyllisyys</li> <li>• maine</li> <li>• tietoturvallisuus</li> <li>• oikea-aikaisuus</li> <li>• ymmärrettävyys</li> </ul> <p>Datan laadun mittaaminen suoritetaan haastatteleamalla kohdeyrityksen työntekijöitä, jotka omassa toimenkuvassaan työskentelevät tutkimusdatan parissa sen elinkaaren eri vaiheissa. Lomakkeen väitteisiin ei kuitenkaan kerätä numeerisia arvosanoja vaan sanallista palautetta.</p>
Haastateltavien valinta	Tunnistetaan työntekijät, jotka vastaavat edellisessä vaiheessa määritettyihin kriteereihin.
Tiedon kerääminen	Haastattelut pidetään ja nauhoitetaan.
Tulosten analysointi	Haastattelutulokset analysoidaan datan laadun ulottuvuuksittain. Jokainen tarkasteltu ulottuvuus käsitellään erikseen ja sen sisällä eri rooleja edustavien työntekijöiden vastauksia verrataan keskenään. Näin pyritään löytämään ulottuvuuksia, joihin liittyy paljon epäkohtia tai joiden sisällä eri rooleilla on eri näkemys sen toteutumisesta.
Kehitystoimenpiteet	Tuloksista tunnistetaan datan laadun osa-alueet/ulottuvuudet, jotka eivät ole hyväksyttäviä. Kehitysehdotuksilla pyritään ottamaan kantaa erityisesti näihin osa-alueisiin/ulottuvuuksiin.

Haastateltavat henkilöt on jaettu kolmeen rooliin: tuottaja, käyttäjä ja ylläpitäjä. Käyttäjä-rooli on lisäksi jaettu kahteen alarooliin: sisäinen käyttäjä ja ulkoinen käyttäjä. Haastateltavien roolit ja vastaavat lukumäärät on esitetty taulukossa 14. Tuottajista yksi toimii kemian analyysidataa tuottavassa laboratoriossa, joka toimittaa datan kohdeyrityksen tietokantaan. Loput kaksi tuottajaa koordinoivat kohdeyrityksen kenttämittaustöitä. Sisäisiä

käyttäjiä on viisi ja ulkoisia käyttäjiä kolme. He käyttävät dataa muun muassa analyysien ja synteesien tekemiseen. Sisäiset käyttäjät lisäksi tarkistavat ja hyväksyvät tuotetun datan. Ulkoiset käyttäjät puolestaan toimittavat omat synteesinsä sisäisille käyttäjille. Ulkoiset käyttäjät ovat kohdeorganisaation ulkopuolella toimivia asiantuntijoita. Tässä työssä tarkasteltavalla datalla on vain yksi ylläpitäjä, jonka vastuualueelle kuuluu sekä laboratorioissa että kentällä tuotetun datan hallitseminen.

**Taulukko 14.** Haastateltavien roolit ja lukumäärät

Rooli	Lukumäärä
Tuottaja	3
Käyttäjä	
-Sisäinen käyttäjä	5
-Ulkoinen käyttäjä	3
Ylläpitäjä	1
Yhteensä	12

Haastatteluissa saadut tulokset analysoidaan datan laadun ulottuvuuksittain. Jokainen tarkasteltu ulottuvuus käsitellään erikseen ja sen sisällä eri rooleja edustavien työntekijöiden vastauksia verrataan keskenään. Näin pyritään löytämään ulottuvuuksia, joihin liittyy paljon epäkohtia tai joiden sisällä eri rooleilla on eri näkemys sen toteutumisesta. Koska arvioinnissa käytetty mittari on haastateltavien näkemykset datan laadusta, on laadun arviointi täten subjektiivista.

Lopuksi tulosten analysoinnin perusteella tunnistetaan osa-alueet tai ulottuvuudet, joiden toteutuminen ei ole kohdeyrityksen mielestä hyväksyttävää tai riittävää. Erityisesti näihin osa-alueisiin tai ulottuvuuksiin pyritään ottamaan kantaa kehitysehdotuksissa. Kehitysehdotukset muodostetaan sekä laadun arvioinnin tulosten että datan hallinnan arvioinnin perusteella.

Tässä työssä käytetty datan laadun kehityksen strategia on prosessilähtöinen. Tavoitteena on muokata ja kehittää kohdeyrityksen prosesseja ja käytäntöjä tehokkaammiksi. Datan tuotantoprosesseja muuttamalla ja kehittämällä voidaan myös parantaa datan laatua. Tämän proaktiivisen lähestymistavan avulla pyritään lisäksi takaamaan vastaisuudessa tuotettavan tutkimusdatan hyvä laatu.

Arvioinnissa tarkasteltu datatyyppe on rakenteellinen data. Tietokannassa sekä taulukkolaskentaohjelmien taulukoissa sijaitseva tutkimustieto on hyvin numeerista, joten se voidaan nähdä rakenteellisena datana. Tarkasteltu tietojärjestelmätyyppi on tutkimusdatan pitkäaikaista säilytystä varten luotu SQL-tietokanta, johon liittyy myös loppukäyttäjälle tarkoitettu verkkopohjainen käyttöliittymä datan hakemista ja muokkaamista varten.

## 5.2 Datan hallinnan arviointi

Datan hallinnan arvioinnissa pyritään löytämään ongelmia ja haasteita liittyen kohdeyrityksen datan hallintaan. Datan hallinnan ongelmia pyritään selvittämään haastattelemalla samoja henkilöitä kuin datan laadun arvioinnissa. Haastateltavilta kerätään vapaamuotoista palautetta datan hallintaan liittyen liitteen A kysymysten mukaan. Osa-alueita, joihin keskitytään ovat muun muassa vastuut ja roolit, prosessit, käytännöt, standardit ja tietojärjestelmät.

Haastattelut pidetään puolistrukturoidusti. Haastateltavilta kysytään liitteessä A esitetyt kysymykset, joiden lisäksi heiltä voidaan kysyä tarkentavia tai laajentavia kysymyksiä tarpeen mukaan. Haastateltavat vastaavat kysymyksiin siltä osin kuin he oman kokemuksensa mukaan osaavat. Tässä työssä käytetty datan hallinnan arvioinnin prosessi on kuvattu taulukossa 15. Datan hallinnan arviointi tapahtuu samaan aikaan datan laadun arvioinnin kanssa. Haastateltavat henkilöt ovat samat kuin datan laadun arvioinnissa.

***Taulukko 15.** Datan hallinnan arvioinnin vaiheet*

Vaihe	Kuvaus
Haastattelukysymysten muodostaminen	Pyritään ottamaan kantaa kuvassa 6 esitettyihin kokonaisuuksiin (organisaatio, käytännöt ja prosessit, standardit, teknologia). Tavoitteena on löytää juurisyyt ongelmiin.
Haastateltavien valinta	Haastateltavat tunnistetaan datan laadun arvioinnin prosessissa.
Tiedon kerääminen	Haastattelut pidetään ja nauhoitetaan. Haastattelu yhdistetään datan laadun arvioinnin haastatteluun.
Tulosten analysointi	Tulokset kerätään analysointivaiheessa POSMAD-viitekehykseen. Näin koko datan elinkaari tulee huomioiduksi. POSMAD myös mahdollistaa elinkaaren vaiheiden sekä datan hallinnan komponenttien (data, prosessit, ihmiset ja organisaatiot, teknologia) välisten vuorovaikutuksien tarkastelun.
Nykytilan määrittely	Määritellään kohdeyrityksen datan hallinnan nykytaso eri kypsyysmallien avulla.
Tavoitetilan määrittely	Määritellään kohdeyrityksen datan hallinnan tavoitetaso ja yksityiskohtaiset datan hallinnan kehittämisen tavoitteet eri kypsyysmallien avulla.
Kehitysehdotusten luominen	Datan laadun ja datan hallinnan arviointien pohjalta luodaan kehitysehdotukset teorialukujen avulla. Datan hallinnan nykyinen taso sekä tavoitetaso ohjaavat kehitysehdotusten luomista. Kehitysehdotuksilla pyritään ottamaan kantaa ongelmien juurisyihin. Kehitysehdotukset toimivat työkaluna kohdeyritykselle tavoitetasoon pääsemisessä.

Haastatteluissa saadut tulokset kootaan analysointivaiheessa POSMAD-viitekehykseen. Näin tuloksia voidaan analysoida niin, että koko datan elinkaari ja kaikki organisaatiotasot tulevat huomioiduksi. Kehyksen avulla datan hallinnan ongelmista saadaan kattava kuva. Kehys myös auttaa datan hallinnan nykytason määrittämisessä.



Power (2011, s. 32) ehdottaa käyttämään jotakin datan hallinnoinnin kypsyysmallia nykytilanteen arvioimiseen sekä tulevaisuuden tavoitetilan kuvailemiseen. Nykytilanteen selvittäminen ja tavoitetilan määrittelemineen auttavat sopivien kehitysehdotusten luomisessa. Tavoitetilan määrittelyssä tulee myös huomioida käytössä olevat resurssit. Jos datan hallinnan kehittämiseen ei ole mahdollista osoittaa resursseja, on turha tavoitella suurta kasvua datan hallinnan kypsyudessa. Tavoitetilan avulla voidaan lisäksi asettaa realistiset odotukset datan hallinnan kehittymiselle.

Tärkeää on myös määritellä jokin aikajänne, jossa tavoitetilaan tulee päästä. Aikajänteen määrittelyssä tulee olla huolellinen; liian lyhyt aikajänne voi olla mahdoton toteuttaa, mikä johtaa helposti pettymykseen. Liian pitkä aikajänne voi taas aiheuttaa kehityksen pysähtymisen. Toiminnan jatkuva kehittäminen voidaan osaltaan varmistaa onnistumiselämysten ja positiivisten tulosten kautta (Niemi & Kontra 2012, s. 4). Tavoitetilan ja aikajänteen tulee tämän vuoksi olla mahdollisimman realistiset, mutta ei liian vaatimatomat.

## 6. TULOKSET

Tässä luvussa esitellään tulokset, jotka saatiin datan laadun sekä datan hallinnan arvioinnissa. Kohdeyrityksessä tutkimusdataa säilytetään sen pitkäaikaiseen tallettamiseen tarkoitettussa tietokannassa. Tutkimusdatan luonnetta on kuvailtu tarkemmin luvussa 1.1. Käyttäjillä on tunnukset tähän tietokantaan, ja he voivat hakea sieltä dataa. Dataa tuotetaan sekä laboratorioissa (näytteiden analysointi) että kenttämittauksilla (esimerkiksi erilaiset vesimittaukset). Laboratoriot joko syöttävät datan itse tietokantaan tai data toimitetaan kohdeyritykselle syötettäväksi. Kentällä suoritettavien mittausten data siirtyy joko automaattisesti tietokantaan reaaliaikaisesti tai pienellä viiveellä suoraan mittalaitteista (loggereista) tai se haetaan mittalaitteilta ja syötetään käsin.

Datan laadun ulottuvuuksien toteutuminen kohdeyrityksessä on kuvattu kootusti taulukossa 16. Arvosanat perustuvat eri roolien kollektiiviseen näkemykseen kustakin ulottuvuudesta.

**Taulukko 16.** Laadun ulottuvuuksien toteutuminen kohdeyrityksessä

Ulottuvuus	Arvosana (hyvä/kohtalainen/huono)
Saatavuus	Kohtalainen
Sopiva määrä	Hyvä
Luotettavuus	Hyvä
Täydellisyys	Kohtalainen
Tiivis esitystapa	Hyvä
Johdonmukainen esitystapa	Kohtalainen
Helppokäyttöisyys	Huono
Oikeellisuus/tarkkuus	Hyvä
Tulkittavuus	Hyvä
Objektiivisuus	Hyvä
Relevanssi/hyödyllisyys	Hyvä
Maine	Hyvä
Tietoturvallisuus	Hyvä
Oikea-aikaisuus	Kohtalainen
Ymmärrettävyys	Hyvä

Alaluvussa 6.1 esitelty tulokset ovat haastateltujen henkilöiden näkemyksiä. Jokaista ulottuvuutta käsitellään kolmen roolin (tuottaja, käyttäjä, ylläpitäjä) näkökulmasta. Tulokset on esitelty hyvin yksityiskohtaisesti, minkä vuoksi alaluvussa 6.2 on tiivistetty tulosten tärkeimmät huomiot sekä eri roolien näkemykset ulottuvuuksien toteutumisesta. Alaluvussa 6.3 on puolestaan esitelty datan hallinnan arviointiin liittyvät tulokset.

## 6.1 Datan laadun arvioinnin tulokset

### Saatavuus

#### Tuottaja

Data on helposti ja nopeasti saatavissa, jos tietää, mistä sen löytää. Välillä datan saamiseksi joutuu tekemään paljon töitä, sillä kollegoilta voi joutua kysymään, mistä etsitty data löytyy. Tällöin datan saatavuus riippuu siitä, ehtivätkö muut auttaa datan löytämisessä.

Aina ei voi myöskään olla varma, löytyykö jokin data tietokannasta. Esimerkiksi uusien henkilöiden perehdyttäminen on hankalaa, sillä ei ole yhtenäistä paikkaa, jonne kaikki data tallennetaan. Verkkolevyä ja dokumenttienhallintajärjestelmää käytetään vielä paljon datan tallettamiseen. Lisäksi datan palauttamista vaikeuttaa se, jos data häviää loggerista laitevian johdosta ennen kuin se on ehditty hakea talteen.

#### Käyttäjä

Myös käyttäjien näkemyksen mukaan data on helposti ja nopeasti saatavissa, jos tietää, mistä se löytyy. Jos dataa ei ole itse ollut mukana tuottamassa, voi sen löytäminen olla haastavaa. Saatavuutta kuitenkin heikentää se, että on harvoin tiedossa, milloin jokin tietty data on tietokannassa saatavissa. Tällöin kyseisen datan perään joutuu kyselemään. Lisäksi dataa on useissa eri paikoissa. Jotain dataa, esimerkiksi joitain alkuaineanalyysijä, ei ole edes vielä tietokannassa, jolloin niiden saatavuus on huono. Niin sanottu offline-tilukko, jonne kerran vuorokaudessa konsolidoidaan eri tuloksia yhdeksi ladattavaksi Access-tilustoksi, on parantanut datan saatavuutta huomattavasti.

Saatavuus myös riippuu paljon siitä, kuinka tottuneesta tietokannan käyttäjästä on kyse. Toisin sanoen, tietokannan käytettävyyttä voi joidenkin kohdalla heikentää datan saatavuutta. Myös kenttähenkilöstön tuottama tuottodata, eli huomiot liittyen mittauksiin, ovat vaikeasti saatavissa.

Vuosia vanhoja dataa voi olla hankala palauttaa, sillä alkuperäisiä tulosraportteja voi olla vaikea löytää. Alihankkijoilla olevan datan palautettavuus riippuu kyseisen yrityksen tietokannan palautuksen toteutuksesta. Kuitenkin kohdeyrityksellä olevan datan palautettavuus saattaa kärsiä, jos datan tuottanut henkilö lähtee yrityksestä. Alkuperäistä dataa ei välttämättä tässä tapauksessa löydetä.

#### Ylläpitäjä

Datan saatavuutta on yritetty parantaa muun muassa offline-tilukon avulla. Yksittäisten datojen saaminen tietokannasta on helppoa, mutta kattavan datapakettin saaminen on puo-

lestaan haastavaa. Jos offline-taulukon data riittää käyttäjälle, on data myös nopeasti saatavissa. Muutosten tekeminen offline-taulukkoon on kuitenkin hidasta ja hyvin resurssi-riippuvaista työtä.

Tietokannasta tehdään päivittäiset varmistukset, joten datan palauttaminen on mahdollista. Palautuksen kuitenkin suorittaa järjestelmän toimittaja, mikä hidastaa datan palautettavuutta. Datan palautettavuuden kannalta on kuitenkin olennaista, että kaikki data on tietokannassa.

## **Sopiva määrä**

### **Tuottaja**

Yhden näkemyksen mukaan datan sopivassa määrässä on kyse siitä, että kuinka helposti tietokannasta saa itselleen sopivan määrän dataa. Tietokannassa on paljon eri dataa, ja jokainen tarvitsee kaikesta datasta vain osan. Kuitenkin datan hallinnan näkökulmasta dataa tuntuu olevan joskus liikaa. Muiden töiden ohella datan laatuun ei ehdi panostaa riittävästi. Datan määrä nähdään sopivaksi käyttäjien tarpeita ajatellen.

### **Käyttäjä**

Yhden näkemyksen mukaan dataa ei voi periaatteellisesti olla koskaan liikaa, kun tutkitaan jotain ilmiötä. Dataa on kuitenkin joka tapauksessa valtavasti. Datan suuri määrä saattaa joskus vaikeuttaa jonkin tietyn datan saatavuutta. Lisäksi datan suuri määrä vaikeuttaa datan tarkistamista ja hyväksymistä, sillä käyttäjät eivät ehdi tarkistaa kaikkea dataa. Jotkut näkevät datan määrän kuitenkin optimaaliseksi, sillä dataa tuotetaan juuri sen verran, kuin mitä monitorointiohjelmassa on määritelty.

Yhden käyttäjän mukaan datan sopivassa määrässä on kyse siitä, miten helposti datan saa karsittua itselleen sopivaksi.

### **Ylläpitäjä**

Dataa on todella paljon, mikä on tietysti ymmärrettävää ottaen huomioon yrityksen toimiala. On varmasti vaikeaa tehdä päätös siitä, milloin dataa on tarpeeksi ja milloin ollaan tutkittu tarpeeksi.

## **Luotettavuus**

### **Tuottaja**

Datan luotettavuus on pääosin hyvä. Datasta löytyvät virheet voivat heikentää datan luotettavuutta, ja virheet voidaan löytää ja arvioida usein vasta jälkikäteen. Lisäksi datan luotettavuuteen vaikuttavat mittauslaitteen sopivuus tehtyyn mittaukseen, mittauksen

suunnittelun huolellisuus sekä kiire, väsymys ja stressi. Ja koska työ on henkilöstöriippuvaista, voi osaamista mittausten suorittamisesta hävitä yrityksestä poistuvan henkilöstön mukana.

Jos mittauksen suorittavat samat henkilöt kuin aiemmin, ovat tulokset yleensä luotettavia. Luotettavuuden todentaminen on kuitenkin nopeaa, jolloin toistomittaukset voidaan tarvittaessa aloittaa nopeasti.

### **Käyttäjä**

Data nähdään pääosin luotettavaksi. Datan luotettavuusepäkohdat liittyvät yleensä näytteenottoon, hydrologisiin olosuhteisiin sekä itse näytteen edustavuuteen. Datan käyttäjät kuitenkin suhtautuvat dataan terveellisen kriittisesti. Virheet datassa on yleensä helppo huomata, sillä ne erottuvat selkeästi muusta datasta.

Yhden näkemyksen mukaan vanhemman datan luotettavuus kärsii, koska uudemmat mittaukset suoritetaan kehittyneemmillä tekniikoilla. Kuitenkin toisen näkemyksen mukaan vanhempi data on taas luotettavampaa, koska se on käynyt läpi useamman iteraatiokierroksen.

### **Ylläpitäjä**

Data on pääosin luotettavaa. Automaatiomittauksiin voi kuitenkin liittyä loggerien tekemiä virheitä. Ongelmakohdat löytyvät kuitenkin helposti, sillä referenssidataa on todella pitkältä aikaväliltä. Lisäksi inhimilliset virheet voivat heikentää datan luotettavuutta.

## **Täydellisyys**

### **Tuottaja**

Vaatimukset datan täydellisyydelle tulevat käyttäjiltä. Datassa pyritään siihen, että se sisältäisi kaikki käyttäjien tarvitsemat arvot. Joitain tarvittavia arvoja kuitenkin puuttuu.

### **Käyttäjä**

Yhden näkemyksen mukaan ei voi olla koskaan tarpeeksi arvoja, sillä aina halutaan uusia analysointimahdollisuuksia. Tärkeimmät parametrit on kuitenkin pystytty määrittelemään.

Datasta kuitenkin puuttuu oleellisia tietoja. Esimerkiksi hydrologian osalta offline-taulukosta puuttuu tarvittavia arvoja. Lisäksi joitain arvoja mitataan vain harvoista näytteistä, jolloin koko parametrin mittaaminen on helppo kyseenalaistaa. Tietokantaan päätyy vain onnistuneiden näytteenottojen data, mutta joillain käyttäjillä on tarve saada tieto myös niistä näytteenotoista, jotka eivät ole syystä tai toisesta onnistuneet. Tällä hetkellä tämä tieto täytyy hankkia pyytämällä.

## Ylläpitäjä

Käyttäjät määrittelevät, mitä parametreja mitataan. Järjestelmän näkökulmasta huomataan nopeasti, jos datasta puuttuu arvoja. Käytännössä tällöin data ei siirry tietokantaan oikein, jolloin järjestelmä ilmoittaa virheestä syöttödatassa. Tietokannan näkökulmasta data sisältää kaikki arvot.

## Tiivis esitystapa

### Tuottaja

Data on pääosin tiiviisti esitetty. Datan tiiviiseen esitystapaan pyritään esimerkiksi selkeillä lomakkeilla ja taulukoilla.

### Käyttäjä

Tietokannasta saattaa joutua hakemaan olennaisia tietoja eri paikoista. Lisäksi offline-tilaukossa on tyhjiä sarakkeita tietyille alkuaineille, joita ei ole lainkaan mitattu. Peruskemian tulokset ja isotooppitulokset samalle näytteelle ovat kahdella eri rivillä. Niiden yhdistäminen samalle riville parantaisi taulukon tiiviyyttä. Offline-tilaukko on myös jaettu kahdeksi taulukoksi, jolloin taulukot joudutaan yhdistämään. Ulkopuolisilta toimijoilta tuleva data on tiivistä, sillä datan esitystapa on sovittu sopimusvaiheessa.

## Ylläpitäjä

Data on pääosin esitetty tiiviisti, mutta joitain taulukoita pitää transponoida käyttäjälle toisin päin. Tietokannan kannalta on helpompaa, jos jokainen analyysi on oma rivinsä. Näin uusia analyysejä on helppo lisätä vain lisäämällä uusia rivejä.

## Johdonmukainen esitystapa

### Tuottaja

Data on suurimmaksi osaksi esitetty johdonmukaisesti. Kuitenkin esimerkiksi offline-tilaukossa sarakkeet ovat tietystä kohtaa alkaen aakkosjärjestyksessä, kun taas alun sarakkeet eivät ole. Taulukosta tulee siis etsiä kohta, josta lähtien sarakkeet ovat aakkosjärjestyksessä. Mittausdatan lisäksi on muuta tietoa, jota ei ole pystytty viemään tietokantaan ja joka on erilaisissa taulukoissa.

### Käyttäjä

Tietokanta sanelee, missä muodossa tulokset viedään tietokantaan. Data on pääosin esitetty johdonmukaisesti. Kuitenkin esimerkiksi kaas- ja vesikemiatatan yhdistäminen on vaikeaa, sillä ne ovat eri muodossa, vaikka kyseessä olisi sama näyte. Lisäksi käyttäjien käyttämät Excel-tilaukot sekä tulostaulukot ovat eri muodossa. Niihin tulee perehtyä ennen kuin niitä voi käyttää.

## Ylläpitäjä

Tietokanta määrittää, missä muodossa dataa voi viedä kantaan. Offline-taulukko on muuttunut jonkin verran vuosien varrella, mutta nyt sen muoto on toistaiseksi vakiintunut.

## Helppokäyttöisyys

### Tuottaja

Jos data haetaan tietokannasta, ei sen käyttäminen ole helppoa ilman Excel-makroja. Hydrogeokemian Excel-makrot ovat auttaneet datojen helppokäyttöisyydessä ja yhdistämisessä. Datan yhdistäminen on siitä huolimatta vaikeaa.

### Käyttäjä

Data vaatii paljon työtä, jotta sitä voi käyttää haluamallaan tavalla. Eri datojen, esimerkiksi mikrobi- ja kaasudatojen yhdistäminen kemian datan kanssa on hankalaa. Tämä mainittiin selkeäksi tarpeeksi. Tuloksien yhdistäminen geologisen datan kanssa olisi myös todella vaikeaa, sillä geologinen data on tallennettu erilaiseen muotoon. Offline-taulukon datan yhdistämisessä muun datan kanssa voi olla ongelmia, sillä sarakkeita on offline-taulukossa niin monta. Lisäksi eri lähteistä tulevissa datoissa on eri muotoilut, jolloin niitä joutuu muokkaamaan Excelissä. Kaiken kaikkiaan tulosten yhdistäminen koetaan todella hankalaksi. Koska muualta tullut data on eri muodossa, ei sitä ole resursipulan johdosta saatu syötettyä tietokantaan.

## Ylläpitäjä

Offline-taulukosta ulospäin dataa on helppo muokata. Datan muokkaaminen tietokannassa on kuitenkin hankalaa. Dataa on syötetty pian kymmenen vuotta tietyssä muodossa, jolloin kannan muokkaaminen ei ole helppoa. Kannasta hakemisen tapoja voidaan kuitenkin muokata.

On suunniteltu, että tietokantaan tulisi jonkinlainen rajapinta, jolla pystyttäisiin itse tekemään raporttikyselyjä kantaan. Resurssien puutteesta johtuen tästä ollaan kuitenkin luovuttu. Muitakin edistyneitä toiminnallisuuksia on esitetty, mutta ne ovat jääneet toteuttamatta. Jos järjestelmässä haluaa yhdistää dataa muiden tutkimusten kanssa, se ei ole helppoa, mutta kuitenkin mahdollista.

## Oikeellisuus/tarkkuus

### Tuottaja

Data on riittävän virheetöntä. Sitä mukaa, kun dataa syötetään tietokantaan, käyttäjät tarkistavat sen. Inhimillisten virheiden määrä pyritään minimoimaan. Myös mittalaitteet voivat aiheuttaa virheitä. Jos mitataan yrityksen itse rakentamilla laitteilla ensimmäistä

kertaa maailmassa, ei ole olemassa referenssidataa, johon tuotettua dataa voitaisiin verrata. On vaikea sanoa, miltä datan pitäisi näyttää. Toisaalta onko valittu menetelmä sopiva kyseiseen mittaukseen?

Jos jokin laite tuottaa virheellistä dataa, sen huomaamiseen voi mennä vuosia. Virheellistä dataa on monelta vuodelta, ja epäselväksi on jäänyt, millä laajuudella.

### **Käyttäjä**

Data on riittävän tarkkaa ja virheetöntä, vaikka välillä siitä voi löytyä inhimillisiä virheitä. Jos esimerkiksi näytteet ovat menneet sekaisin, tulevat virheet todella helposti esille. Pienet virheet on vaikeampi havaita. Datan virheettömyys kuitenkin nähdään korjaantuvan ajan suhteen. Yhden näkemyksen mukaan vastuu datan virheettömyydestä on sekä datan tuottajilla että käyttäjillä.

### **Ylläpitäjä**

Data on riittävän virheetöntä. Tämän ulottuvuuden toteutumisessa ovat tarkastus- ja hyväksymistoimenpiteet isossa roolissa. Datan käyttäjät ovat kuitenkin kiireisiä, joten tämä tarkastus-hyväksymisprosessi tietokannassa voi helposti jäädä väliin. Tutkijoilla voi olla valtava määrä tutkimuksia ja kova kiire, jolloin he eivät ehdi panostaa datan tarkastamiseen. Se voi vaikuttaa datan oikeellisuuteen. Joskus virheet huomataan vasta, kun dataa raportoidaan ja siitä piirretään kuvaajia.

## **Tulkittavuus**

### **Tuottaja**

Datan mittausyksiköt ovat selkeät ja dataa on helppo tulkita. Makroja käytetään muokkaamaan laitteilta tulevat yksiköt käyttäjille sopiviksi.

### **Käyttäjä**

Pääosin datan mittausyksiköt ovat selkeät ja dataa on helppo tulkita. Kuitenkin joissain Excel-taulukoissa mittausyksiköt eivät ole selkeitä, mutta yksikkö silti selviää aina jostain. Tietokannasta otettua dataa on helppo tulkita. Joissain taulukoissa on kuitenkin epäjohdonmukaisuuksia. Taulukolla voi olla useampi tekijä, jolloin samoista näytteenotto-paikoista voidaan käyttää eri nimityksiä.

### **Ylläpitäjä**

Datan mittausyksiköt ovat selkeät.



## **Objektiivisuus**

### **Tuottaja**

Data on objektiivisesti kerättyä. Työohjeilla, selkeillä lomakkeilla ja perehdyttämällä pyritään siihen, että asioita tehdään mahdollisimman samalla tavalla. Jotkin mittaukset kuitenkin perustuvat datan kerääjän omaan arviointiin. Lisäksi mittaajahenkilöt pyritään pitämään samoina mahdollisimman pitkään, jotta arvioinnin ovat johdonmukaisia.

### **Käyttäjä**

Data on kerätty objektiivisesti. Näytteitä ei ole jätetty ottamatta siinä pelossa, että paljastuisi joitain ikäviä tietoja. Näytteenottojen priorisointi on kuitenkin osittain subjektiivista, ja kenttäarvot ovat ihmisten keräämiä dataa. Näitä ei kuitenkaan koeta ongelmaksi. Joissain tapauksissa monitorointiohjelma sanelee, milloin ja missä näytteenottoja suoritetaan.

### **Ylläpitäjä**

Data on kerätty objektiivisesti.

## **Relevanssi/hyödyllisyys**

### **Tuottaja**

Data on hyödyllistä työlle ja se soveltuu myös hyvin työhön. Datasta joutuu kuitenkin joskus tekemään omia koosteita tai dataa täytyy muokata paljon. Vanhaa dataa käytetään hyödyksi uusissa mittauksissa samoissa kohteissa.

### **Käyttäjä**

Data koetaan hyödylliseksi ja soveltuu hyvin työhön.

### **Ylläpitäjä**

Datan käyttäjät ovat todennäköisesti muokanneet datan sellaiseksi, että he saavat parhaan hyödyn siitä. Tietokantahakuja on muokattu käyttäjien näkökulmaa ajatellen.

## **Maine**

### **Tuottaja**

Datalla on hyvä maine ja data tulee hyvistä lähteistä. Datan mainetta voi kuitenkin heikentää muun muassa se, että mittauslaite on vuosia tuottanut väärää dataa.

## Käyttäjä

Datalla on hyvä maine. Sen tuottamiseen käytetään hyviä ja luotettavia laboratorioita. Datasta ollaan myös saatu kansainvälisesti hyvät arviot. Dataa hyödynnetään ja julkaitaan kuitenkin tarpeettoman vähän, ottaen huomioon sen hyvä maine.

## Ylläpitäjä

Virhetilanteet voivat heikentää datan mainetta. Loggerien tuottamassa datassa on ollut suuriakin virheitä. Tällainen väistämättä vaikuttaa datan maineeseen.

## Tietoturvallisuus

### Tuottaja

Laboratorion LIMS-järjestelmään (Laboratory Information Management System) pääsee vain sisäverkosta. Tietokone vaatii käyttäjätunnuksen ja salasanan, samoin järjestelmä. Datan lähettäminen sähköpostitse on vähentynyt. Dataan pääsyä on rajoitettu sopivasti.

### Käyttäjä

Dataan pääsyä on rajoitettu sopivasti. Käyttäjillä oleva data on suojattu kyseisen henkilön ja hänen yrityksensä tietoturvakriteerien mukaisesti. Tietokannassa katkeavat istunnot haittaavat joskus datan hakemista. Datan ei kuitenkaan koeta olevan salaista, sillä useimmiten se julkaistaan myöhemmin julkisesti joka tapauksessa.

### Ylläpitäjä

Data on suojattu riittävän hyvin luvattomalta pääsylvä. Ei voida kuitenkaan olla varmoja siitä, mitä ihmiset tekevät heidän tietokoneillaan olevan datan kanssa. Datan käyttäjät allekirjoittavat luottamuksellisuussitoumuksen ja heille tehdään suppea turvallisuustarkastus ennen kuin heille annetaan pääsy sisäverkkoon. Lisäksi ulkopuolisten datan käyttäjien käyttämä tietokanta on peilikanta varsinaisesta tuotantokannasta, johon ei pääse ulkoverkosta. Dataan pääsyä on pyritty rajoittamaan sopivasti kuitenkin sillä ajatuksella, että toiminnan halutaan olevan mahdollisimman avointa.

## Oikea-aikaisuus

### Tuottaja

Menetelmiä viedä data tietokantaan pitäisi kehittää, sillä nyt se on liian vaikeaa tai liian aikaavievää. Lisäksi on paljon syviä kairanreikiä ja tulppavälejä, joista ei ole otettu näytteitä vuosiin resurssipulan takia. Tämä on ongelma monitoroinnin kannalta, sillä kemian puolella muutoksia voi tapahtua lyhyessä ajassa.

## Käyttäjä

Peruskemian tulokset ovat riittävän oikea-aikaisia, mutta isotooppituloksissa ja erityislaboratorioista tulevissa tuloksissa kestää kauemmin; joissain isotooppituloksissa voi kestää useita kuukausia. Ne eivät välttämättä ehdi valmistua nykyisissä raportointiaikatauluissa. Joitain tuloksia myös pyydetään suoraan laboratoriolta ennen kuin ne ovat tietokannassa. Datassa olisi myös hyvä olla tieto siitä, mitä siitä puuttuu ja mitä on vielä tulossa. Joidenkin käyttäjien työssä myös vanhalla datalla on suuri merkitys.

## Ylläpitäjä

Laboratorio toimittaa datat melko lyhyessä ajassa. Loggerien tuottamassa datassa on joskus laitteista johtuvia ongelmatilanteita. Tämä vaatii taas ylläpitäjän osalta paljon käsi-työtä, jotta kyseisen datan saa siirrettyä tietokantaan. Kun ulkoisilta toimijoilta on tilattu tutkimuksia, on datan voinut saada nopeammin sähköpostitse. Nyt kuitenkin ollaan määritelty sopimuksissa, että työ on tehty, kun data on kohdeyrityksen tietokannassa.

## Ymmärrettävyys

### Tuottaja

Dataa on helppo ymmärtää, mutta datan merkitys on joskus vaikeampi ymmärtää. Datan merkitys, toisin sanoen miksi dataa tuotetaan, on loppukäyttäjille ja osalle tuottajista selvä. Kuitenkin esimerkiksi kenttähenkilöstölle datan merkitys ei ole aina selvä, mikä saattaa vaikuttaa motivaatioon tai johtaa datan keräämisen kyseenalaistamiseen.

### Käyttäjä

Käyttäjät kokevat datan ymmärrettävyyden hyväksi. Data on kuitenkin kompleksista ja se käsittelee monimutkaisia järjestelmiä, joten datan ymmärtäminen vaatii asiantuntijuutta. Sisäankoulutus datan ymmärtämiseen tulee kollegoilta ja työyhteisöltä. Datan ymmärrettävyyttä voidaan kuitenkin parantaa ja helpottaa työkaluilla. Jos esimerkiksi offline-taulukkoa ei olisi, moni asia jäisi tekemättä.

### Ylläpitäjä

On melko selvää, mitä ja miksi ollaan mitattu. Datan keräämisellä ja sen analysoinnilla on kuitenkin todella iso merkitys koko yrityksen toiminnalle ja projektille.

## 6.2 Laadun arvioinnin tulosten tiivistys

Tässä luvussa tiivistetään datan laadun arvioinnin tulokset ja eri roolien näkemykset samoista ulottuvuuksista. Lisäksi tuloksista nostetaan esille tärkeimmät huomiot, mutta tuloksista ei kuitenkaan tehdä suurempia päätelmiä. Datan laatu nähdään pääosin hyväksi.

Kuitenkin datan helppokäyttöisyydessä on ongelmia. Lieviä ongelmia on lisäksi datan saatavuudessa, täydellisyydessä, johdonmukaisuudessa ja oikea-aikaisuudessa.

### Saatavuus

Datan saatavuus koetaan pääosin hyväksi. Kuitenkin datan useat talletuspaikat vaikeuttavat datan saatavuutta. Jotain dataa ei myöskään ole vielä viety tietokantaan. Saatavuus riippuu myös siitä, kuinka tottuneesta tietokannan käyttäjästä on kyse. Käyttäjien mielestä offline-tilukko on parantanut data saatavuutta. Kenttähenkilöstön tuottaman tuottodatan sekä vanhojen tuloraporttien saatavuus nähdään huonoksi. Datan palautettavuus on varmistettu, jos data sijaitsee tietokannassa.

### Sopiva määrä

Datan määrä yleisesti ottaen on valtava. Sopivan määrän nähdään liittyvän siihen, kuinka helposti datamäärän saa karsittua itselleen sopivaksi, sillä datan valtava määrä koetaan tarpeelliseksi. Datan määrä kuitenkin asettaa haasteita datan hallinnalle.

### Luotettavuus

Kaikki roolit kokevat datan luotettavuuden pääosin hyväksi. Inhimilliset virheet sekä loggerien tuottamat virheet kuitenkin heikentävät datan luotettavuutta. Virheiden havaitseminen koetaan helpoksi. Näkemykset siitä, miten datan luotettavuus kehittyy ajan myötä, eroavat hieman käyttäjien kesken. Näkemykset eivät ole kuitenkaan toisensa poissulkevia.

### Täydellisyys

Tuottajat näkevät, että vaatimukset datan täydellisyydelle tulevat käyttäjiltä. Tuottajat ja käyttäjät näkevät kuitenkin joitain puutteita datan täydellisyydessä. Tietokannan näkökulmasta datan täydellisyys on hyvä.

### Tiivis esitystapa

Tuottajat näkevät datan esitystavan tiiviiksi. Kuitenkin käyttäjät kokevat, että offline-tilukko voisi olla muotoiltu tiiviimmin. Joitain datan esitystavan ratkaisuja on jouduttu tekemään tietokannan ehdoilla.

### Johdonmukainen esitystapa

Datan esitystavan johdonmukaisuus koetaan pääosin hyväksi. Tietokanta sanelee, missä muodossa dataa voi syöttää kantaan. Tämän johdosta kuitenkin jotain dataa ei ole pystytty viemään vielä tietokantaan. Joitain dataa on vaikea yhdistää, sillä niiden esitystapa on eri, vaikka kyseessä olisi sama näyte. Excel-tilukkojen ja tulostaulukkojen johdonmukaisuus vaihtelee.

## Helppokäyttöisyys

Datan helppokäyttöisyys nähdään todella heikoksi. Datan muokkaamiseen sopivaan muotoon käytetään erilaisia työkaluja, esimerkiksi Excel-makroja. Eri datojen yhdistäminen on myös erittäin hankalaa. Datan muokkaaminen myös tietokannassa on vaikeaa. Tietokannan käytettävyyttä ollaan pohdittu parannettavan erilaisilla käyttöliittymäratkaisuilla, mutta niitä ei ole toteutettu.

## Oikeellisuus/tarkkuus

Kaikki roolit kokevat datan riittävän virheettömäksi. Datasta kuitenkin tiedetään löytyvän inhimillisiä virheitä, minkä lisäksi laitteet voivat tuottaa virheellistä dataa pitkällä aikavälillä. Datan virheettömyyden nähdään paranevan ajan funktiona. Datan tarkastus- ja hyväksymisprosessilla on suuri merkitys tämän ulottuvuuden toteutumisella, mutta käyttäjien kiireet haittaavat datan huolellista tarkastamista.

## Tulkittavuus

Datan tulkittavuus koettiin pääosin hyväksi. Kuitenkin joidenkin datojen epä johdonmukaisuuksien mainittiin haittaavan tulkittavuutta.

## Objektiivisuus

Kaikki roolit näkevät, että data on kerätty objektiivisesti.

## Relevanssi/hyödyllisyys

Data koetaan hyödylliseksi ja se soveltuu käyttäjien työhön. Tämä vaatii kuitenkin datan muokkausta.

## Maine

Tuottajat ja käyttäjät kokevat datan maineen erittäin hyväksi. Ylläpidon mukaan esimerkiksi loggerien tuottama virheellinen data heikentää datan mainetta.

## Tietoturvallisuus

Kaikki roolit kokevat, että dataan pääsyä on rajoitettu sopivasti. Datan tietoturvallisuuden on kiinnitetty huomiota, ja käyttäjien tiedot tarkastetaan ennen kuin heille annetaan pääsy dataan.

## Oikea-aikaisuus

Tuottajat näkevät, että hankala datan vieminen tietokantaan heikentää datan oikea-aikaisuutta. Lisäksi joitain näytteitä ei ole otettu resurssipulan takia, jolloin kyseinen data puut-

tuu. Käyttäjien mukaan peruskemian tulokset ovat riittävän oikea-aikaisia, mutta isotooppituloksien valmistumisessa voi kestää useita kuukausia. Ylläpidon mukaan loggerien tuottaman datan manuaalinen korjaaminen heikentää kyseisen datan oikea-aikaisuutta. Datan toimitustavan määrittäminen sopimuksissa alihankkijoiden kanssa on parantanut uuden datan oikea-aikaisuutta.

### Ymmärrettävyys

Tuottajien mukaan datan merkitys ei ole aina kaikille sidosryhmille selvä, mikä saattaa johtaa toiminnan kyseenalaistamiseen. Käyttäjät kokevat datan ymmärrettävyyden hyväksi, vaikka data on todella kompleksista. Ylläpito ymmärtää datan merkityksen.

## 6.3 Datan hallinnan arvioinnin tulokset

Tässä alaluvussa esitellään tulokset datan hallinnan haastattelukysymyksille. Tulokset on lajiteltu haastateltavien roolien mukaan. Esitellyt tulokset ovat haastateltujen henkilöiden näkemyksiä.

### Tuottaja

Laboratorion käyttämä LIMS-järjestelmä on todella vanha, mikä osaltaan haittaa datan käytettävyyttä. Datan siirtäminen LIMS:stä tietokantaan vaatii makrojen käyttämistä, ja datan muokkaaminen tietokannasta käytettävään muotoon vaatii taas makrojen käyttämistä. Tietokannasta myös huomaa, että se on täysin räätälöity järjestelmä. Se on kuitenkin toiminut viime aikoina paremmin kuin aluksi, mutta se ei ole pysynyt kohdeyrityksen kehityksen mukana, kun on tullut uusia tapoja tarkastella dataa. Excel-makrot ovat kuitenkin auttaneet datan käyttämisessä.

Datan tuottajilla jatkuva kiire vaikuttaa keskittymiseen ja työmotivaatioon, kun asioita ei useinkaan saa valmiiksi. Tuottajien toive on mahdollisuus tehdä asiat huolella ja ajan kanssa. Tuottajilla on myös paljon erityyppistä dataa, joka pitäisi saada tietokantaan. Ja koska on paljon erityyppistä dataa, on vaikea muistaa, minne mitään dataa tallennetaan. Joitain taulukoita verkkolevyllä voi jäädä täyttämättä, koska niitä on niin monta. Varsinaisen mittausdata on melko helppo viedä, mutta erilaiset huomiot voivat jäädä merkitsemättä. Hydrologian puolella on tärkeää, että kaikki on kirjattu ylös. Tuottajat toivovat selkeää roolitusta siihen, kenen vastuulla minkäkin datan vieminen tietokantaan on. Nyt tuottajilla on tunne, että asioita oletetaan liikaa ja joskus jopa unohdetaan.

Koska toiminta on jokseenkin kaaosmaista, on uusien työntekijöiden perehdyttäminen hankalaa. Myös kenttähenkilöstön motivointi liittyy paljolti siihen, miksi asioita ja tiettyjä mittauksia tehdään. Datan keräysmenetelmiä on niin monta, että niiden kaikkien hallitseminen on haastavaa. Tietokannassa on lisäksi tarkistamatonta dataa, jota kuitenkin käytetään raporteissa. Aikataulusyistä näin kuitenkin joudutaan tekemään.

Tuottajat toivoisivat apua tietokoneohjelmien käyttöön sekä datan tallettamiseen. Lisäksi kun mittauksia ja projekteja suunnitellaan, pitäisi datan hallinnan ammattilainen saada projektiin jo tässä vaiheessa mukaan. Nyt on käynyt niin, että dataa ollaan kerätty, mutta sitä ei ole viety tietokantaan. Myös jos kohdeyritys aikoo laajentaa toimintaansa ja alkaa myydä asiantuntijapalveluitaan asiakkailleen, tulisi käyttäjäroolitus tietokannassa toteuttaa niin, ettei kaikilla ole oikeuksia asiakkaiden datoihin.

Datan käyttäminen voi tapahtua vasta vuoden datan keräämisen jälkeen. Jos käyttäjät kyselevät tuottajilta datan keräämiseen liittyviä yksityiskohtia näin pitkän ajan jälkeen, on tuottajien vaikea muistaa kyseisiä yksityiskohtia. Lisäksi kommunikointia tehtävistä mittauksista pitäisi kehittää. Nyt on vaarana, että samoja mittauksia toistetaan, kun kaikki eivät tiedä suoritetuista/suoritettavista mittauksista. Lisäksi tuotettua dataa ei osata mahdollisesti hyödyntää muissa projekteissa, kun siitä ei olla tietoisia.

### Käyttäjä

Tietokanta nähdään enemmän datan säilytyspaikkana, sillä se ei ole joustava datan käyttäjien tarpeita ajatellen. Eri tieteenaloja on vaikea vetää yhteen. Tietokantaa tulisi kehittää käytettävyyden näkökulmasta. Dataa on myös valtavasti, joten on usein tilanteita, joissa täytyy tietää, keneltä kysyä. Käyttäjille ei ole selvää, missä kaikki data on. Datan suuri määrä vaikeuttaa myös oikean, raportoidun datan löytämistä. Tietokannassa pitäisi myös tuoda selkeästi esille mittausarvot, joita ei tulisi käyttää, sillä niitä on joskus päätynyt raportteihin. Koska dataa on paljon, ei sen tarkistaminen voi olla kovin seikkaperäistä. Ulkoiset käyttäjät voisivat mahdollisesti tarkistaa dataa, sillä he ehtisivät käyttää tarkistukseen enemmän aikaa. Jotkin käyttäjät myös kokevat tarkistamisen työlääksi.

Hydrologian puolella taas ulkoisilla käyttäjillä on paljon dataa, joka ei ole helposti sisäisten käyttäjien saatavissa. Ja koska mikrobi- ja kaasudatalle ei ole offline-taulukkoa vastaavaa ratkaisua, joutuvat sisäiset käyttäjät pyytämään ulkoisilta käyttäjiltä kyseistä dataa helpommin käytettävässä muodossa.

Data pitäisi saada helposti muokattavaan muotoon. Työkalut ovat käyttäjille tärkeitä, ja ne tulisi liittää kemian lisäksi mikrobi- ja kaasudataan. Tämän tulisi myös mahdollistaa eri datojen yhdistämisen. Lisäksi työkaluja ja makroja tulisi jatkuvasti kehittää, kun käyttäjien tarpeet muuttuvat. Jos työkalujen aktiivista ylläpitoa ja käytön jatkuvuuden turvaamista ei varmisteta, ongelmat tulkintapuolella lisääntyvät. Tämä nähtiin vakavana huolenaiheena.

Kun ulkoiset käyttäjät havaitsevat virheitä datassa, he lähettävät siitä sähköpostia sisäisille käyttäjille. Tähän toivotaan selkeää protokollaa, miten toimia, kun virhe havaitaan. Tämä voisi olla esimerkiksi jokin muodollinen kaavake, jolla asia ilmoitetaan sekä selkeä käytäntö, minne kaavake toimitetaan. Lisäksi käyttäjille tulisi ilmoittaa, kun virhe on korjattu. Tämä on kaikki ollut sähköpostin varassa, mikä koetaan huonoksi tavaksi hoitaa

tällaisia asioita. Itse dataan pitäisi myös tulla merkintä, että sitä korjataan. Näin kolmannet osapuolet saavat myös tiedon datan korjaamisesta.

Käyttäjien toimintaa myös haittaa vakava resurssiongelma. Varsinaista työtä ei ole ehditty tekemään, sillä uusien työkalujen opettelu ja projektointi vievät paljon aikaa. Tämä näkyy ongelmana raportoinnissa, sillä joissain raporteissa on usean vuoden julkaisuviive. Resurssiongelma vaikuttaa myös virheiden korjaukseen. Virheitä ei ole ehditty korjaamaan, jolloin niitä päättyy todennäköisemmin raportteihin. Käyttäjät joutuvat tekemään kompromisseja laadun suhteen pysyäkseen aikatauluissa. Joitain alkuainetuloksia monen vuoden ajalta ei ole viety tietokantaan, koska tähän ei ole ollut aikaa tai henkilöstöresursseja. Käyttäjät toivovat riittäviä henkilöstöresursseja tai rahallisia resursseja ulkopuolisen työvoiman käyttöön. Datan tuottamiseen ja tallentamiseen tulisi panostaa, jotta kaikki yksityiskohdat ovat vielä pitkän ajan päästä saatavissa ja tulkittavissa.

Datan tuottaminen nähdään sirpalemaiseksi. Kaikkea dataan liittyvää tulisi koordinoida hallitusti. Datan hallinnalle tulisi myös asettaa selkeät roolit, vastuut ja protokollat. Sekä datan tuottajille että käyttäjille pitäisi antaa aikaa pohtia datan hallinnan toteutusta. Sähköpostikirjeenvaihtoa pitäisi saada vähennettyä. Tämä helpottaisi myös datan korjaajan työtä, sillä kaikki ilmoitukset tulevat samassa muodossa.

Näytteiden analyysijä teetetään olosuhteiden pakosta eri laboratorioissa, joten tuloksia saadaan eri formaateissa. Tämä muun maussa vaikeuttaa datan viemistä tietokantaan. Sopimuksissa tulisi mahdollisesti määritellä jokin yhteinen formaatti, jossa data toimitetaan. Tietokantaan ei myöskään ole mahdollista tallentaa tietoa tai yksityiskohtia käytetystä analyysimenetelmästä. Tämä tieto pitäisi pystyä linkittämään dataan vielä vuosikymmenien jälkeenkin. Haasteena nähdään se, miten data pysyy käyttökelpoisena koko projektin ajan.

## Ylläpitäjä

Dataa tuotetaan ja käytetään melko pienellä henkilöstöllä, mutta silti jokaisella on omat tarpeensa. Tämä on hankalaa ylläpitämisen näkökulmasta. Lisäksi jokaisella tuntuu olevan kova kiire. Se aika, joka voidaan yleensäkin käyttää dataan, on liian pieni. Henkilöstöä myös lähtee organisaatiosta enemmän kuin uutta saadaan tilalle. Tämä väistämättä johtaa aikataulujen venymiseen. Töitä on muutenkin todella paljon.

Datan syöttäminen tietokantaan ei ole kenties aivan niin helppoa kuin se voisi olla. Jos dataa ei syötä tietokantaan säännöllisesti, voi syöttäminen tuntua vaikealta. Tietokannan kehittämiseen liittyen on muutenkin ollut toive visuaalisesta käyttöliittymästä. Lisäksi käyttäjät haluavat enemmän kaiken datan, kuin yksittäisiä paketteja. Tähän on pyritty keksimään jonkinlaista ratkaisua.



Datan ylläpidon kannalta nykyinen järjestelmän on toimiva, mutta loppukäyttäjien näkökulmasta se on vaikeakäyttöinen. Tietokannan päälle olisi melko helppo kehittää helppokäyttöisempi käyttöliittymä. On aikoinaan suunniteltu, että tietokannan päälle tulisi jonkinlainen liiketoimintatiedon raportoinnin kaltainen rajapinta. Resurssien puutteesta johtuen tämä on jäänyt. Tietojärjestelmän uudistukset voivat tuoda lisäarvoa käyttäjille, mutta muutosten taloudellisuus on myös huomioitava. Järjestelmäkehityksessä tulee business case määritellä hyvin, sillä resurssit ovat niin vähäiset.

Kenttämittausten tekniikassa ja prosessissa nähdään myös olevan kehitettävää. Koko prosessiin ja käytettyyn tekniikkaan, esimerkiksi loggerien toimintaan, on mahdollista lisätä älykkyyttä. Järjestelmään, joka vastaa datan siirtämisestä loggereista tietokantaan, ollaan saamassa tarvittavia muutoksia. Vaikka järjestelmä toimii hyvin, on sen hallittavuus hankala. Kyseessä on räätälöity ohjelmisto, jonka toimittajalla on yksi järjestelmän ylläpitoon kykenevä henkilö. Yhden pienen muutoksen tekemiseen saattaa mennä vuosia aikaa.

Tieto hankittavasta datasta tulee ylläpidolla melko myöhään. Kun mittauksia suunnitellaan, pitäisi avainhenkilöt ottaa jo tässä vaiheessa mukaan mittaukseen. Kommunikointia voisi mahdollisesti muutenkin kehittää. Jos käyttäjät kyselevät kenttähenkilöstöltä vuotta aiemmin kerätystä datasta, ei kenttähenkilöstö välttämättä muista, mitä silloin on tapahtunut.

Lisäksi ulkopuolisia toimijoita käytetään paljon mittausten tekemiseen, jolloin dataa tulee eri muodoissa ja eri tavalla. Vaikka kaikilla on luonnollisesti eri tarpeet, olisi hyvä määritellä jonkinlaiset määrittelyt tai ääriviivat, joiden puitteissa toimitaan.

## 7. TULOSTEN POHDINTA JA KEHITYSEHDOTUKSET

Tässä luvussa esitellään empirian ja teorian pohjalta tehtyjä päätelmiä. Datan hallinnan ongelmat ja haasteet kootaan POSMAD-viitekehykseen. Näin tuloksista voidaan tehdä yksityiskohtaisempia johtopäätöksiä kattaen koko datan elinkaari ja organisaation eri komponentit. Viitekehykseen lajiteltujen ongelmien lisäksi luvussa pohditaan myös muita datan hallinnan ongelmia.

Lisäksi luvussa määritellään kohdeyritykselle datan hallinnan tämän hetkinen kypsyystaso sekä tavoitetaso, jota organisaation tulisi tavoitella. Tavoitetason määrittelyssä otetaan huomioon se, kuinka voimakkaasti kohdeyritys voi realistisesti keskittyä datan hallintaan ja sen kehittämiseen. Luvun lopuksi esitellään kehitysehdotukset havaittujen ongelmien korjaamiseksi.

### 7.1 POSMAD-viitekehys

Liitteessä D on esitetty datan hallinnassa havaitut ongelmat ja haasteet sijoitettuna POSMAD-matriisiin. Datan hävittämistä ei tarkastella, sillä kohdeorganisaatiossa tuotettua dataa ei lähtökohtaisesti hävitetä missään vaiheessa. Matriisissa ei ole esitettynä joitain tuloksissa esille tulleita ongelmia, sillä näitä ei voida sijoittaa sujuvasti kyseiseen taulukkoon. Nämä ongelmat käsitellään myöhemmin tässä luvussa.

Taulukosta voidaan huomata, että ongelmat painottuvat datan elinkaaren kohdalla lievästi käyttövaiheeseen. Tämä tarkoittaa, että datan käyttäjät kohtaavat suuren osan datan hallinnan ongelmista omassa työssään. Vaikka ongelmien juurisyyt johtaisivat muualle, ongelmat näkyvät selkeimmin käyttäjille. Ongelmia on kuitenkin melko tasaisesti kaikissa tarkastelluissa datan elinkaaren vaiheissa. Organisaation komponentteja tarkasteltaessa ongelmat jakautuvat myös melko tasaisesti. Seuraavaksi POSMAD-matriisiin lajiteltuja ongelmia esitellään tarkemmin organisaatiokomponentteittain.

#### Data

Suunnitteluvaiheessa ei ole määritelty selkeää formaattia tai muotoa, jossa toimittajan tulisi toimittaa data kohdeyritykselle. Tämän vuoksi eri laboratoriot toimittavat dataa eri muodoissa, jolloin kohdeyritykselle on kertynyt paljon eri formaatissa olevaa dataa, jota ei ole voitu syöttää tietokantaan. Dataa ei ole myöskään ehditty syöttämään tietokantaan osittain resurssipulasta johtuen. Tietokantaan vietyyn dataan ei voida myöskään linkittää yksityiskohtia laboratorion käyttämästä analyysimenetelmästä. Ylläpidolle datan valtava

määrä taas asettaa hallinnallisia ja ylläpidollisia haasteita. Käyttäjien dataan liittyvät ongelmat johtuvat pääosin laatuepäkohdista. Lisäksi datan sijainti ei ole aina käyttäjille selvä.

## Prosessit

Suunnitteluvaiheessa datan viemistä tietokantaan ei ole huomioitu riittävän hyvin. Suunnittelemassa tarkasti ja ottamalla huomioon datan vieminen jo suunnitteluvaiheessa voidaan varmistaa datan helppo tallettaminen. Myös datan keräysmenetelmiä (esimerkiksi loggereista) on monia, mikä on haaste datan tuottajille. Dataa myös toimitetaan kohdeyritykselle eri tavoilla. Kenttämittausten prosessiin olisi ylläpidon mukaan mahdollista lisätä automaatiota/tekoälyä, mikä taas vähentäisi manuaalityön määrää. Datassa olevien virheiden korjaamiselle ei ole myöskään määritelty prosessia tai protokollaa. Virheiden ilmoittaminen tehdään tällä hetkellä sähköpostin avulla. Virheellisen datan korjaaminen saattaa myös olla aikaa vievää ja työlästä. Käyttäjät kokevat datan tarkastamisen ja hyväksymisen työlääksi. Tämä on vakava ongelma, sillä datan tarkastaminen ja hyväksyminen koetaan tärkeäksi datan virheettömyyden varmistamiseksi. Tämä myös johtaa siihen, että tarkastamatonta dataa käytetään raporteissa. Datan käyttö saattaa myös tapahtua vasta vuosia varsinaisen mittauksen jälkeen, jolloin datan tuottamiseen liittyvien yksityiskohdien saaminen on vaikeaa.

## Ihmiset ja organisaatiot

Suunnitteluvaiheessa datan hallinnan ammattilaista ei ole otettu riittävän ajoissa mukaan mittaukseen, mikä on saattanut johtaa ongelmiin datan tallettamisvaiheessa. Tämä on myös edesauttanut datan hallinnan siilomaisuutta. Datan hallinta toimii pieninä henkilöryhminä, jolloin koko yrityksen datan hallinnan toimivuus riippuu paljon näiden siilojen välisestä kommunikoinnista. Tulosten mukaan kommunikointia tehtävistä mittauksista tulisikin kehittää. Datan merkitys ei ole myöskään selvä kaikille datan keräämisestä vastaaville henkilöille. Tämä saattaa johtaa ongelmiin motivaatiossa, mikä voi taas johtaa huolimattomuuteen. Myös uusien työntekijöiden perehdyttäminen on hankalaa datan hallinnan kaaosmaisuuksien vuoksi. Tämä voi johtua myös osaltaan siitä, että roolit ja vastuut eri datojen viemisestä tietokantaan eivät ole selkeät.

Kohdeyrityksen resurssipula näkyy selkeimmin ihmiset ja organisaatiot -komponentissa, ja se aiheuttaa useita muita ongelmia. Dataa ei ole ehditty viedä tietokantaan henkilöstöpulan vuoksi. Myös datan ylläpidon henkilöstö on pieni, mutta kuitenkin jokaisella käyttäjällä on hieman omanlaisensa tarpeet. Niiden täyttäminen on tämän vuoksi haastavaa. Käyttövaiheessa resurssiongelma johtaa raportoinnin viivästymiseen, virheiden korjaamisen viivästymiseen sekä kompromisseihin laadussa.

## Teknologia

Teknologiaan liittyvät ongelmat keräämisvaiheessa liittyvät pääosin loggerien toimintaan. Niiden toiminta koetaan epävarmaksi ja niistä kerätään dataa eri menetelmillä. Lisäksi nähdään, että loggerien toimintaan voitaisiin lisätä automaatiota tai tekoälyä manuaalituksen vähentämiseksi. Datan useat talletuspaikat sekä datan viemisen hankaluus vaikeuttavat datan tallettamista.

Myös ylläpidon näkökulmasta tietokantaan liittyy ongelmia. Datan muokkaaminen tietokannassa on hyvin vaikeaa. Datan esitystapaa voidaan käytännössä muuttaa vain muokkaamalla hakutoimintoja. Tietokannan kehitys on kuitenkin järjestelmätoimittajan vastuulla, joten muutosten saaminen on toimittajariippuvaista. Tietokannan käyttäjäkokemuksen kehittäminen on myös lähestulkoon mahdotonta resurssipulan takia. Tämän vuoksi tietokannan käyttö koetaan vaikeaksi talletus- ja käyttövaiheessa. Myös loggeridatan tietokantaan siirtämisestä vastaavan järjestelmän ylläpito ja kehittäminen on hankalaa, sillä muutosten saaminen on vahvasti toimittajariippuvaista.

Käyttäjät näkevät, että datan käyttöä helpottavien työkalujen ja makrojen jatkuva kehitys on varmistettava, ja kehityksen tulisi tapahtua yhdessä käyttäjien kanssa. Tämä nähdään kuitenkin tällä hetkellä haasteena, mutta mahdollisena tulevaisuuden ongelmana.

## 7.2 Muut huomiot tuloksista

Kohdeyrityksen datan hallinta on syntynyt aikoinaan liiketoimintatarpeesta. Haastattelujen mukaan datan tallettaminen ja säilöminen on kehittynyt pelkistä Excel-taulukoista vuosien varrella 90-luvulta lähtien laajaksi tietokannaksi. Datan hallinta on noussut organisaatiossa enemmän tai vähemmän alhaalta ylös (bottom-up), mikä on saattanut johtaa datan hallinnan siilomaisuuteen. Datan hallinta, suunnittelusta käyttöön, suoritetaan hyvin pitkälti pieninä siiloina tai henkilöryhminä. Koko organisaation datan hallinnan toimivuus on tällöin hyvin riippuvaista näiden siilojen välisestä kommunikaatiosta.

Lisäksi kohdeyrityksen toiminta on hyvin keskittynyttä ydinliiketoimintaan. Osittain tämä saattaa johtua koko organisaatiota vaivaavasta resurssiongelma. Tällöin kuitenkin ydinliiketoimintaa tukevat tukitoiminnot, kuten datan hallinta, kärsivät. Tukitoimintojen puutteellisuus taas heijastuu varsinaisen liiketoiminnan sujuvuuteen. Resurssiongelma vaikuttaa tulosten mukaan koko organisaation toimintaan. Kohdeyritys joutuu tasapainottelemaan resurssien allokoinnissa ydinliiketoiminnan ja tukitoimintojen välillä. Resurssiongelma johtuu osittain kohdeyrityksen omistajien asettamista kustannuspaineista alan huonon markkinatilanteen vuoksi.

Kohdeyrityksen vähäisiä tukitoimintoja selittää myös se, että yrityksen tietohallinto toimii konsernitasolla. Tämä tarkoittaa, ettei tietohallinnolla ole välttämättä tarkkaa ymmärrystä kohdeyrityksen liiketoimintatarpeista. Kohdeyrityksellä ei ole tällöin myöskään

mahdollisuutta allokoida tietohallinnon resursseja itselleen parhaiten sopivalla tavalla. Lisäksi ulkoistetussa tietohallinnossa kohdeyrityksen datan hallintaan liittyviä asioita hoitaa yksi henkilö osa-aikaisesti.

Kohdeyritys aikoo myös tulevaisuudessa laajentaa toimintaansa palveluliiketoimintaan ja myydä asiantuntijaosaamistaan tuleville asiakkailleen. Palveluliiketoimintaan siirtymisen aiheuttaa uusia haasteita datan hallinnalle. Jos kohdeyrityksen tarjoamat palvelut vaativat datan hallintaa, tulee datan hallinnan prosessien olla riittävän tehokkaita kannattavan toiminnan takaamiseksi. Yleisestikin tehottomat datan hallinnan prosessit, kuten mitkä tahansa tehottomat liiketoimintaprosessit, aiheuttavat yritykselle turhia kustannuksia. Jos näitä kustannuksia ei voida helposti kohdistaa datan hallintaan, voivat kustannukset näkyä muissa toiminnoissa. Tällöin niitä ei välttämättä osata tarkastella datan hallinnasta johtuvina kustannuksina.

### **7.3 Datan hallinnan nykytaso**

Kohdeyrityksen datan hallinnan kypsyysden nyky- ja tavoitetasot määritellään käyttäen neljää eri kypsyysmallia. Tavoitetason määrittelyssä otetaan huomioon kohdeyrityksen rajalliset resurssit. Liian suurta kehitystä datan hallinnan kypsyysdessä ei ole mielekästä lähteä tavoittelemaan hyvin rajallisilla resursseilla. Tavoitetasoksi pyritään määrittelemään taso, jonka yritys pystyy realistisesti saavuttamaan nykyisellä resurssitasolla tai pienillä resurssilisäyksillä.

Datan hallinnan kypsyysden nykytason määrittelyyn käytetään luvussa 3.1 esiteltyjä Brennanin (2012), IBM:n (2007) ja COBIT:n (ITGI 2007) kypsyysmalleja sekä taulukossa 11 esiteltyä Loshinin (2011) datan hallinnan käytäntöjen kypsyysmallia. Tarkasteluun on valittu kyseiset mallit, sillä ne sopivat parhaiten kohdeyrityksen tapaukseen, ja niillä voidaan parhaiten kuvata kohdeyrityksen tilannetta.

Taulukossa 17 on esitetty kootusti kohdeyrityksen sijoittuminen kussakin mallissa. Brennanin (2012) mallin mukaan kohdeyrityksen nykyinen taso on kaoottinen. Yrityksen lähestymistapa datan hallintaan ja sen ongelmiin on reaktiivinen eikä sillä ole datan hallinnan strategiaa. Nämä ovat osa kaoottisen tason määrittelyä. Kuitenkin esimerkiksi yrityksen käyttämä tietokanta asettaa joitakin standardeja datalle ja sen laadulle koskien lähinnä datan muotoilua ja täydellisyyttä.

**Taulukko 17.** Kohdeyrityksen datan hallinnan kypsyyden nykytaso (arvioitu taso korostettuna)

Malli	Tasot					
Brennan (2012)	<b>Kaoottinen</b>	Reaktiivinen	Määritelty	Proaktiivinen	Ennakoiva	
IBM (2007)	<b>Taso 1</b>	Taso 2	Taso 3	Taso 4	Taso 5	
COBIT (ITGI 2007)	0 Ei datan hallintaa	<b>1 Varhainen/Ad hoc</b>	<b>2 Toistettava mutta intuitiivinen</b>	3 Määritelty	4 Johdettu ja mitattava	5 Optimoitu
Loshin (2011)	<b>Varhainen</b>	<b>Toistettava</b>	Määritelty	Hallittu	Optimoitu	

IBM:n (2007) mallin mukaan kohdeyrityksen datan hallinnan nykytaso on taso 1. Mallissa käytettävän käsitteen projekti voi kohdeyrityksen kontekstissa muuttua mittaukseksi. Mallin toinen taso vaatii yritykseltä proaktiivista lähestymistapaa datan hallintaan. Lisäksi metadata nähdään tällä tasolla osana kriittisten dataelementtien dokumentointia. Tulosten mukaan kohdeyrityksessä esimerkiksi tietoa käytetystä analyysimenetelmästä, joka on metadataa, ei voida helposti linkittää dataan. Ongelmien ratkaisu kohdeyrityksessä on myös reaktiivista. Näistä syistä johtuen kohdeyritys ei yllä tasolle 2.

COBIT:n (ITGI 2007, s. 144) kypsyyksimallissa kohdeyritys on tasolla 1 tai 2, riippuen siitä, tiedostaako koko kohdeyrityksen organisaatio tarpeen tehokkaalle datan hallinnalle ja siitä, voidaanko datan ylläpitäminen nähdä synonyyminä datan hallinnalle. Datan ylläpito on osoitettu tietohallinnon avainhenkilölle, mutta tämä ei välttämättä tarkoita, että myös datan hallinta olisi selkeästi osoitettu. Taso 3 vaatii muun muassa datan hallinnan suorituskyky mittareita, joita kohdeyrityksellä ei ole käytössä.

Loshinin (2011, s. 47) datan hallinnan käytäntöjen kypsyyksimallin mukaan kohdeyrityksen taso on varhaisen ja toistettavan välissä. Tulosten mukaan henkilöstö tekee toisteisia toimintoja ilman koordinoitua, mutta organisaatio kuitenkin pyrkii yhdistämään datan yhteen lähteeseen. Joitain käytäntöjä, kuten datan tarkastaminen ja hyväksyminen, on määritelty. Joitain käytäntöjä, kuten virheiden korjaaminen, ei ole määritelty.

## 7.4 Datan hallinnan tavoitetaso

Tavoitetason määrittelyssä tulee olla realistinen. Esimerkiksi Brennanin (2012) mallissa proaktiivisen tason saavuttaminen vaatii jo huomattavaa keskittymistä ja investointi datan hallinnoinnin kehittämiseen (Brennan 2012). Kuten luvussa 5.2 on todettu, realististen tavoitteiden avulla voidaan hallita organisaation odotuksia liittyen datan hallinnan kehittämiseen. Lisäksi realistisella kehityksen aikajänteellä voidaan varmistaa onnistumisen mahdollisuus. Taulukossa 18 on esitetty kootusti tavoitetasot datan hallinnan kypsyydelle eri mallien mukaan.

**Taulukko 18.** Kohdeyrityksen datan hallinnan kypsyyden tavoitetasot (tavoitetasot on korostettu)

Malli	Tasot					
Brennan (2012)	Kaoottinen	<b>Reaktiivinen</b>	Määritelty	Proaktiivinen	Ennakoiva	
IBM (2007)	Taso 1	<b>Taso 2</b>	Taso 3	Taso 4	Taso 5	
COBIT (ITGI 2007)	0 Ei datan hallintaa	1 Varhainen/Ad hoc	2 Toistettava mutta intuitiivinen	<b>3 Määritelty</b>	4 Johdettu ja mitattava	5 Optimoitu
Loshin (2011)	Varhainen	<b>Toistettava</b>	<b>Määritelty</b>	Hallittu	Optimoitu	

Työn tekijä on valinnut tavoitetasot kohdeyrityksen resurssimahdollisuuksien sekä 2-3 vuoden aikajänteen perusteella. Toisin sanoen kohdeyrityksen nykyisillä resursseilla tai pienillä resurssilisäyksillä ja noin 2-3 vuoden kehittämishankkeen aikana kohdeyrityksellä on realistiset mahdollisuudet saavuttaa valitut tavoitetasot. Kohdeyrityksen vähäisten resurssien johdosta jokaisessa mallissa pyritään käytännössä yhden tason parannukseen. Eri mallien tavoitetasojen määrittelyjen kohdat toimivat kohdeyrityksen datan hallinnan kehittämisen yksityiskohtaisina tavoitteina. Nämä on koottu taulukkoon 19. Yksityiskohtaiset tavoitteet ovat siis tavoitetason vaatimuksia, joita kohdeyritys ei vielä täytä nykyisellä toiminnallaan.

Brennanin (2012) mallissa reaktiivinen taso vaatii standardien ja olennaisten laadun hallinnan prosessien käyttöönottoa. Näitä on jo kohdeyrityksessä joissain määrin käytössä, mutta niitä tulisi kehittää. Esimerkkinä kehitettävästä standardista on datan toimitusformaatin määrittäminen. Kehitettäviä laadun hallinnan prosesseja ovat esimerkiksi tarkastus-/hyväksymisprosessit sekä virheistä ilmoittaminen ja niiden korjaaminen. Taso vaatii lisäksi datan hallinnan suunnitelman tunnistamisen ja määrittämisen. Tällaista ei ole vielä kohdeyrityksessä määritelty. Suunnitelma datan hallinnalle toimii myös pohjana datan hallinnan jatkuvalla kehittämiselle. Tason mukaan myös strategia datan hallinnalle tulee määritellä ja se tulee kommunikoida. Tämä on tärkeää, sillä datan hallinnan strategian avulla datan ja sen hallinnan merkitys yritykselle voidaan vakiinnuttaa ja kommunikoida koko organisaatiolle.

IBM:n (2007) mallissa toinen taso vaatii, että dataan liittyviä säännöksiä on dokumentoitu enemmän ja että ne on julkaistu koko organisaatiolle. Näitä voivat olla esimerkiksi ylläpidon toivomat ääriiviivat toiminnalle. Ongelmien ratkaisua tulisi tehdä proaktiivisella ja tiimiperusteisella lähestymistavalla sekä toistettavilla prosesseilla. Tämä tarkoittaa, että organisaation tulee yhdessä osata ratkoa ongelmia ennen kuin ne varsinaisesti kehittyvät ongelmiksi. Prosessien tulee olla määriteltyjä ja tehokkaita, jotta ne voivat olla helposti toistettavia. Mallissa myös mainitaan metadata tärkeäksi osaksi kriittisten dataelementtien dokumentointia. Tämä tarkoittaa kohdeyrityksen tapauksessa esimerkiksi sitä, että tieto käytetyistä analyysimenetelmistä pitää pystyä linkittämään varsinaiseen dataan.

**Taulukko 19.** *Datan hallinnan kehityksen tavoitteet*

Malli	Tavoitteet
Brennan (2012)	<ul style="list-style-type: none"> <li>Datan hallinnan ja laadun standardeja on otettu käyttöön.</li> <li>Olennot datan laadun hallinnan prosessit on otettu käyttöön.</li> <li>Datan hallinnan suunnitelma on tunnistettu.</li> <li>Datan hallinnoinnin strategia on määritelty ja kommunikoitu.</li> </ul>
IBM (2007)	<ul style="list-style-type: none"> <li>Dataan liittyviä säännöksiä on dokumentoitu ja julkaistu koko organisaatiolle.</li> <li>Ongelmia ratkaistaan enemmän proaktiivisella ja tiimiperusteisella lähestymistavalla sekä toistettavilla prosesseilla.</li> <li>Metadata on tärkeä osa kriittisten dataelementtien dokumentointia (analyysimenetelmän yksityiskohdat).</li> </ul>
COBIT (ITGI 2007)	<ul style="list-style-type: none"> <li>Datan hallinnan tarve niin tietohallinnossa kuin koko organisaatiossa ymmärretään kautta linjan.</li> <li>Vastuu datan hallinnasta on määritelty.</li> <li>Datan omistajuus on osoitettu vastuussa oleville henkilöille.</li> <li>Datan hallinnan prosessit on määritelty (tietohallinnossa).</li> <li>Datan hallintaa seurataan jossain määrin.</li> <li>Alkeelliset suorituskykymittarit on määritelty.</li> <li>Datan hallinnan henkilöstöä koulutetaan.</li> </ul>
Loshin (2011)	<ul style="list-style-type: none"> <li>Datan hallinnan tavoitteita asetetaan.</li> <li>Datalähteitä arvioidaan (esimerkiksi IPMAP-työkalun avulla).</li> <li>Organisaatio pyrkii (vahvemmin) yhdistämään datan yhteen lähteeseen.</li> <li>Varhaiset käytännöt dataan liittyvien ongelmien ratkomiseksi on määritelty.</li> <li>Datan laadun harjoittajat kaappaavat parhaita käytäntöjä.</li> <li>Datan laadun palvelutasosopimuksia (esimerkiksi datan toimitusformaatti) on määritelty.</li> </ul>

COBIT:n (ITGI 2007) mallissa määritellyn tason mukaan tarve datan hallinnalle on ymmärretty niin tietohallinnossa kuin koko organisaatiossa. Tämä tarkoittaa, että niin tiedonhallinnan asiantuntijoiden kuin datan käyttäjien ja tuottajien tulee ymmärtää datan hallinnan merkitys yritykselle. Myös vastuun datan hallinnasta tulee olla määritelty ja datan omistajuus osoitettu vastuussa oleville henkilöille. Datan hallintaa tulisi seurata ja alkeellisia suorituskykymittareita tulisi olla määritelty (esimerkiksi aikamääreiden sisällä tietokantaan vietyjen datapakettien osuus). Lisäksi mallin mukaan datan hallinnan prosessien tulee olla tietohallinnon määrittelemiä. Käytännössä prosessien määrittelemisen tietohallinnossa voi olla ongelmallista, sillä tietohallinto toimii konsernitasolla, eikä sillä välttämättä ole tarkkaa ymmärrystä kohdeyrityksen prosesseista. Tämän vuoksi datan hallinnan prosessit tulisi määritellä kohdeyrityksessä tiedonhallinnan asiantuntijoiden avustamana. Lopuksi datan hallinnan henkilöstöä tulisi kouluttaa jollain tasolla.



Loshinin (2011) mallissa tavoitteeksi on asetettu taso toistettavan ja määritellyn välillä. Toistettavalla tasolla yrityksen tulisi määritellä varhaiset käytännöt dataan liittyvien ongelmien ratkomiseksi (esimerkiksi virheiden korjaaminen). Määritellyllä tasolla yrityksen tulee luoda datan hallinnan tavoitteita ja datan hallintaa harjoittavat henkilöt kaappaavat parhaita käytäntöjä niin organisaation sisällä kuin sen ulkopuolelta. Yrityksellä tulisi myös olla selkeä prosessi datalähteiden arvioinnille. Lisäksi datan laadun palvelutasosopimuksia on määritelty käytäntöjen seuraamiseksi (esimerkiksi datan toimitusformaatti).

Yhdistämällä eri mallien yksityiskohtaisia tavoitteita voidaan kohdeyrityksen tavoitteet esittää tiivistetysti seuraavasti:

1. Datan hallinnan strategia on määritelty ja kommunikoitu.
2. Tarve datan hallinnalle ymmärretään koko organisaatiossa, ja tietämystä datan hallinnasta levitetään.
3. Olennaiset datan hallinnan prosessit on määritelty ja otettu käyttöön.
4. Datan hallinnan vastuut ja roolit on määritelty, nimetty ja osoitettu selkeästi.
5. Datan hallintaan ja laatuun liittyviä standardeja ja säännöksiä on määritelty, otettu käyttöön ja viestitty koko organisaatiolle.
6. Datan hallinnan alkeellisia suorituskykymittareita on määritelty, ja datan hallintaa seurataan jossain määrin.
7. Ongelmien ratkaisu on proaktiivista, tiimiperusteista ja toistettavaa.
8. Metadata on tärkeä osa kriittisten dataelementtien dokumentointia.
9. Datan hallinnan henkilöstöä koulutetaan.
10. Datan laadun palvelutasosopimuksia on määritelty.
11. Datalähteitä arvioidaan.
12. Dataa pyritään yhdistämään yhteen lähteeseen.
13. Datan hallinnan parhaita käytäntöjä kaapataan.

Datan hallinnan strategiassa tulee määritellä muun muassa mihin datan hallinnalla pyritään, datan hallinnan tavoitteet, keinot, miten tavoitteeseen tullaan pääsemään sekä aikajänne, jossa tavoitteet tulee saavuttaa. Tavoitteena on myös organisaation ajatuksellinen muutos, jonka johdosta datan hallinnan merkitys ymmärretään koko organisaatiossa. Tässä kohtaa esimerkiksi datan hallinnan ja laadun puolestapuhujan rooli korostuu. Melkein kaikki tavoitteet kuitenkin liittyvät pelkistetysti siihen, että kohdeyrityksen datan hallinnan toiminnan tulisi olla määritellympää, ohjatumpaa, tehokkaampaa ja toistettavampaa.

## 7.5 Kehitysehdotukset

Kehitysehdotukset luodaan soveltaen datan hallinnan ja datan laadun kehittämisen teoriaa kohdeyrityksen datan hallinnan ja laadun ongelmiin. Luvussa 7.4 määritellyt datan hal-

linnan tavoitteet ohjaavat kehitysehdotusten luomista niin, että ehdotukset tähtäävät kyseisten tavoitteiden saavuttamiseen. Kehitysehdotukset auttavat siis kohdeyritystä kehittämään tutkimusdatan hallintaa kohti tavoitetasoa. Datan laadun kehittämisen näkökulmasta esitelty kehitysehdotukset perustuvat prosessilähtöiseen strategiaan. Kohdeyrityksen datan laatu koetaan pääosin hyväksi, joten tässä tapauksessa prosessilähtöinen strategia on datalähtöistä soveltuvampi. Prosessilähtöisen strategian avulla voidaan myös ottaa kantaa paremmin datan hallinnallisiin ongelmiin eikä vain datan laadullisiin epäkohtiin.

Kohdeyrityksen datan hallinta on tuloksista päätellen vielä hyvin alkeellista. Tämän vuoksi ei ole mielekästä antaa kohdeyritykselle kovin spesifisiä kehitysehdotuksia, jotka keskittyisivät vain tiettyjen ongelmien korjaamiseen. On hyödyllisempää antaa perustavanlaatuisia kehitysehdotuksia, jotka auttavat muun muassa suunnittelemaan toimintaa tehokkaammaksi ja tekemään datan hallinnasta johdetumpaa.

## 1. Datan hallinnoinnin suunnittelu ja jalkauttaminen

Datan hallinnoinnilla voidaan parantaa datan laatua kannustamalla yritystä käytäntöjen ja prosessien määrittelyyn sekä organisaatioiden väliseen yhteistyöhön (IBM 2007, s. 3). Lisäksi sillä voidaan kehittää yrityksen tietovirtoja ja muuttaa prosesseja johdonmukaisemmiksi (Newman & Logan 2006, s. 4). Datan hallinnoinnin huolellisella suunnittelulla ja jalkauttamisella voidaan ottaa kantaa useaan kohdeyrityksen datan hallinnan ongelmaan: tehottomat ja määrittämättömät prosessit, epäselvät vastuut/roolit, datan hallinnan siilomaisuus ja teknologisten ratkaisujen haasteet.

Kun datan hallinnoinnissa otetaan huomioon datan hallinnoinnin viitekehys, datan elinkaaren vaiheet sekä organisaation eri komponentit, voidaan dataa johtaa kokonaisvaltaisesti. Datan laadun kehittämisessä tuleekin keskittyä niin ihmisiin, prosesseihin kuin teknologiaan, jotta hankkeessa voidaan onnistua (Friedman 2006, s. 4). Osa datan hallinnointia on vastuiden ja roolien määrittely sekä organisaation hallinta niin, että ongelmia voidaan ratkaista proaktiivisesti. Niin sanotusta tulipalojen sammuttamisesta tulee päästä eroon ja ongelmia tai haasteita pitää pystyä ratkomaan jo ennen kuin ne ovat varsinaisesti ongelmia. Datan hallinnoinnin perustamiseen voidaan käyttää esimerkiksi kuvan 8 datan hallinnoinnin perustamisen prosessia. Tämä ohjaa perustamista oikeaan suuntaan ja auttaa huomioimaan tarvittavat näkökulmat. Datan hallinnoinnin perustamisen yhteydessä voidaan myös määritellä ja kommunikoida datan hallinnoinnin strategia.

Kun datan hallinnointia organisoidaan, ei rooleista ja toiminnoista käytetyillä termeillä ole juurikaan merkitystä. Tärkeää on se, miten nämä on määriteltä ja miten ne jalkautetaan. Jos datan hallinnointi voidaan määritellä selkeästi ja jalkauttaa onnistuneesti, on yrityksellä jo hyvät mahdollisuudet tehokkaaseen datan hallinnointiin. (Cohen 2006.) On kuitenkin tärkeää, että asioista puhutaan samalla kielellä; määrittelyjen tulee olla selkeät kaikille organisaatiossa.

## 2. Datavastaavan nimeäminen ja datan isännöinnin jalkauttaminen

Luvuissa 3.2 ja 3.4 käsiteltiin datan isännöintiä, joka on osa datan hallinnointia. Se voidaan kuitenkin nähdä hyvin tärkeänä osana datan hallinnoinnin onnistumista, minkä johdosta se on nostettu erilliseksi kehitysehdotukseksi. Ilman datan isännöintiä voi datan hallinnointi jäädä vain hyvien aikomusten viitekehyyksi, jota ei koskaan jalkauteta operatiivisella tai taktisella tasolla (Plotkin 2014, s. 3). Datavastaavat ja datan isännöitsijät siirtävät korkean tason linjaukset käytäntöön. Tavoitteena on, että strategiset tavoitteet muutetaan operatiivisiksi käytännöiksi, joita noudatetaan taktisella tasolla. Isännöinnillä on tämän vuoksi suuri merkitys datan hallinnoinnin onnistumiselle.

Datavastaavan rooli on muuttaa strategiset linjaukset taktiseksi suunnitelmaksi, jota datan isännöitsijä toteuttaa. Datavastaava vastaa muun muassa muutoshallinnasta, sidosryhmien hallinnasta, käyttäjien esille tuomien ongelmien ratkaisemisesta sekä siitä, että data ja työkalut soveltuvat käyttäjien tarpeisiin. Datan isännöitsijällä on taas yksityiskohtaista tietämystä liiketoimintaprosesseista, käyttäjien vaatimuksista sekä käytetyistä teknologisista ratkaisuista. (Cheong & Chang 2007, s. 1005.) Nämä kaksi roolia täydentävät toisiaan ja yhdessä vastaavat datan isännöinnin kattavasta jalkauttamisesta.

Osa isännöintiä on esimerkiksi datan hallintaan ja datan laatuun liittyvien standardien ja säännöksiä määrittely ja käyttöönotto. Lisäksi datavastaaja ja datan isännöitsijä voivat määritellä datan hallinnan suorituskykykymittareita. Kohdeyhteyksen tulee määritellä sekä datavastaavan ja datan isännöitsijän vastuut ja tehtävät, nimetä kyseiset vastuut sekä selkeästi osoittaa roolit sopiville henkilöille. Tärkeää on, että rooleista vastaavat henkilöt ymmärtävät roolin merkityksen ja heille osoitetut vastuut. Datavastaavan tehtävä on myös implementoida top-down-kontrolli datan hallintaan, jotta kohdeyhteyks voi ottaa kantaa datan hallinnan siilomaisuuteen ja sen asettamiin ongelmiin.

Kohdeyhteyksen tapauksessa datavastaaja ja datan isännöitsijä voivat olla esimerkiksi osa kohdeyhteyksen omassa organisaatiossa sijaitsevaa datan/tiedonhallinnan ryhmää. Konsernitason tietohallinnon vastuut eivät muuttuisi, mutta edellä mainitun ryhmän tehtävä olisi vastata kohdeyhteyksen spesifisiin tietotarpeisiin sekä johtaa datan hallintaa. Tärkeää on, että datan hallinnan asiantuntijat sijaitsevat omassa organisaatiossa. Näin heillä on mahdollisuus ymmärtää riittävän hyvin kohdeyhteyksen liiketoimintatarpeita.

## 3. Datan hallinnoinnin ja laadun kehittämisen kriittisten onnistumistekijöiden tunnistaminen ja parhaiden käytäntöjen huomioiminen

Luvussa 4.2 esiteltiin datan hallinnoinnin ja laadun kehittämisen hankkeiden kriittisiä onnistumistekijöitä. Näiden tekijöiden huomioimisella tärkeiden hankkeiden yhteydessä voi olla merkittävä vaikutus hankkeen onnistumiseen. Kohdeyhteyksen tulee tunnistaa omalle hankkeelleen kriittisimmät onnistumistekijät ja ottaa ne huomioon koko hankkeen ajan.

Johdon tuki todettiin tärkeimmäksi onnistumistekijäksi. Johdon sitouttaminen on kuitenkin haastavaa, joten riittävä johdon tuki tulisi varmistaa ennen hankkeen aloitusta tai suunnitteluvaiheessa. Johdon sitouttaminen vaatii houkuttelevan business casen: hyvällä datan hallinnoinnilla voidaan saada kustannussäästöjä, jotka tekevät toiminnasta kannattavampaa (Niemi & Kontra 2012, s. 1). Tätä tulisi lähestyä datan hallinnan ongelmien liiketoimintavaikutusten kautta, toisin sanoen miten ongelmat vaikuttavat yrityksen liiketoimintaan. Mitä kustannussäästöjä datan hallinnan kehittämisellä voidaan saavuttaa?

Samoin kuin kohdeyrityksen tulisi tunnistaa onnistumistekijät tulisi sen myös huomioida datan hallinnan parhaat käytännöt. Parhaiden käytäntöjen huomioiminen liittyy myös datan hallinnoinnin suunnitteluun ja organisoimiseen. Kaksi tärkeintä parasta käytäntöä liittyen datan hallintaan ovat datan ja sen laadun merkitystä korostava palkitsemisjärjestelmä sekä kokonaisvaltainen lähestyminen datan hallintaan. Kokonaisvaltaisella lähestymisellä tarkoitetaan, että datan hallinnoinnissa todella huomioidaan teknologian lisäksi myös ihmisiin ja prosesseihin.

#### 4. Datan hallinnan prosessien määrittelemine ja kehittäminen

Vaikka prosessien määrittelemine on osa datan hallinnointia ja datan isännöintiä, on sitä silti kohdeyrityksen tapauksessa syytä korostaa. Kuten kaikki tehottomat liiketoimintaprosessit, tehottomat datan hallinnan prosessit aiheuttavat ylimääräisiä kustannuksia yritykselle. Ne myös hidastavat työnteoa ja sitovat liikaa henkilöresursseja. Datan tuottajien ja käyttäjien näkökulmasta tehottomat prosessit voidaan myös kokea turhauttaviksi.

Kohdeyrityksen kohdalla tulisi kiinnittää erityisesti huomiota liitteen D POSMAD-viitekehyksen prosessit-komponentin ongelmiin. Prosessien määrittelyllä ja kehittämisellä voidaan ottaa kantaa muun muassa näihin ongelmiin. Määrittelyä vaativia prosesseja ovat esimerkiksi virheiden korjaus ja datan toimitustapa. Datan toimitustavan määrittelyssä voidaan hyödyntää esimerkiksi tavoitteissa mainittua palvelutasosopimusta. Kehitystä vaativia prosesseja ovat muun muassa datan tarkastaminen/hyväksyminen sekä mittaus-ten suunnittelu.

Kohdeyritys on myös aikeissa laajentaa toimintaansa palveluliiketoimintaan. Jos tulevien palveluiden tarjoaminen asiakkaille vaatii datan hallintaa kohdeyrityksen osalta, on datan hallinnan prosessien oltava riittävän kustannustehokkaita ja toistettavia. Sama koskee oikeastaan kaikkia palveluiden tuottamiseen liittyviä liiketoimintaprosesseja. Kun prosessit ovat mahdollisimman optimoituja, voidaan varmistaa palvelutoiminnan kannattavuus sekä palveluiden korkea laatu.

#### 5. Riittävien resurssien turvaaminen datan hallinnan kehittämiseen ja ylläpitämiseen

Kohdeyritystä vaivaava resurssipula vaikuttaa koko organisaatioon, mutta erityisesti ihmiset ja organisaatio -komponenttiin sekä teknologia-komponenttiin. Resurssipulasta taas juontaa useita muita ongelmia. Jotta kohdeyritys saa esimerkiksi vietyä tietokantaan

kaiken sinne kuuluvan datan, on tälle varmistettava riittävät henkilöresurssit. Lisäksi jos tavoitteena on kehittää mittausautomaatiota ja/tai tietokannan käyttäjäkokemusta, on myös teknologisten ratkaisujen kehittämiseksi varattava riittävät rahalliset sekä henkilöresurssit.

Myös itse datan hallinnan kehittäminen ja ylläpitäminen vaativat resurssien turvaamista. Jos nykyisillä työntekijöillä ei ole aikaa, eikä uusia työntekijöitä palkata, voi datan hallinnan kehittämistä olla vaikea toteuttaa. Esimerkiksi datan hallinnoinnin perustaminen vaatii jonkin verran suunnittelutyötä. Lisäksi jos suunniteltujen palveluiden tuottaminen asiakkaille vaatii datan hallintaa, on laajentuneelle datan hallinnan toiminnalle varattava sen vaatimat lisäresurssit. Tämä tarve on pystyttävä kommunikoimaan tehokkaasti johdolle vakuuttavan business casen avulla: investoinnit datan hallintaan mahdollistavat prosessien ja teknologisten ratkaisujen kehittämisen niin, että kustannuksia voidaan saada pysyvästi nykyistä tasoa alemmas. Tämä taas on liiketoiminnallisesti kannattavaa pitkällä aikavälillä. Riittävillä resursseilla voidaan myös mahdollistaa työn korkea laatu.

On kuitenkin syytä huomioida, että resurssien lisääminen ei ole niin sanottu hopealuoti ongelmien korjaamiseen. Resurssit toki varmistavat toiminnan jatkumisen ja ylläpitämisen sekä mahdollisesti kehittämisen, mutta ne toimivat vain kehityksen mahdollistajana. Resurssien lisäksi on kehityksessä otettava myös muita asioita huomioon ja tehtävä töitä kehityksen eteen. Aiemmat kehitysehdotukset liittyvätkin näihin näkökulmiin.

## 6. Muut ehdotukset

Datan hallinnan kehittäminen ei saa jäädä kertaluonteiseksi hankkeeksi. Esimerkiksi kuvan 8 datan hallinnoinnin perustamisen mallissa viimeisellä vaiheella ei ole selkeää alkua ja loppumispäivää. Datan hallinnoinnin ylläpito ei lopu ennen kuin datan hallinnointi on täysin sisäistetty organisaatioon (Ladley 2012, s. 57). Ylläpidolla myös varmistetaan, että kehityshankkeen tavoitteet saavutetaan. Toisaalta datan hallinnoinnin voi myös nähdä jatkuvan kehityksen näkökulmasta kuten kuvassa 7. Tarkoituksena on, ettei organisaatio taannu tai tyydy nykytasoon, vaan pyrkii aina kehittämään toimintaansa.

Kohdeyrityksen ei tulisi myöskään lähteä kehittämään datan laatua datalähtöisesti. Uutta dataa kertyy niin nopeasti, että olemassa olevaa dataa korjaamalla on vaikea kehittää kaiken kerättävän/kerätyn datan laatua. Prosessilähtöisellä datan laadun kehittämisellä otetaan kantaa dataa tuottaviin prosesseihin niin, että ne tuottavat laadukkaampaa dataa. Kun nämä prosessit ovat kunnossa, voidaan harkita olemassa olevan datan korjaamista.

Lisäksi kohdeyrityksen tulee levittää tietoisuutta ja ymmärrystä datan hallinnasta sekä sen tarpeesta. Kun koko organisaatio ymmärtää datan hallinnan tarpeen ja merkityksen, voidaan sen asema organisaation toiminnassa vakiinnuttaa. Datan hallinnan puolestapuhujilla (*champion*) on suuri merkitys datan hallinnan sanoman levittämisessä. Hänen tehtä-

vänä on kannustaa organisaatiota ja johtoa sitoutumaan hankkeeseen. Tämä onnistuu esimerkiksi perustelemalla datan hallinnan liiketoiminnallinen arvo ja kommunikoimalla selkeä business case johdolle ja koko organisaatiolle.

Lisäksi on varmistettava, että tietohallinto ymmärtää kohdeyrityksen tiedolliset liiketoimintatarpeet. Tietohallinnon perimmäinen tehtävä on tuottaa sen asiakkaille niiden tarvitsemat tietopalvelut. Tämän tekee kohdeyrityksen tapauksessa haastavaksi se, että tietohallinto toimii konsernitasolla. Kirjallisuus korostaa liiketoiminnan ja tietohallinnon yhteistoimintaa (Panian 2010, s. 944; Page 2011, s. 18; Malange et al. 2015, s. 1). Tämän vuoksi liiketoiminnan ja tietohallinnon on tärkeää kehittää tietopalveluita yhdessä sellaisiksi, jotka palvelevat liiketoimintatarpeita ja ovat samalla tehokkaita ja kannattavia.

Datan tallettamiseen käytetty tietokanta on olennainen työkalu melkein kaikissa datan elinkaaren vaiheissa. Datan tallettamisen yhteydessä tulee datan viemisen olla riittävän helppoa, jotta kaikki data voidaan mahdollisimman pian viedä tietokantaan. Näin data on nopeasti ja helposti kaikkien saatavilla. Datan käytön osalta tietokannassa tulee keskittyä käyttäjäkokemuksen kehittämiseen. Kohdeyrityksessä on jo pohdittu hyviä vaihtoehtoja käytettävyyden parantamiseksi, mutta näitä ei ole toteutettu. Parhaimmassa tapauksessa tietokanta kannustaa käyttäjiään perehtymään dataan.

Datan hallinnan siilomaisuus asettaa haasteita kommunikoinnille. Kohdeyrityksen tulee kehittää keinoja, joiden avulla kaikki sidosryhmät saavat olennaiset tiedot esimerkiksi uusista datapaketeista, tulevista mittauksista, tehtävistä/tehdystä korjauksista sekä muutoksista käytäntöihin ja linjauksiin. Lisäksi datan, siihen liittyvien prosessien ja työvaiheiden merkityksen tulee olla selvä kaikille, jotka ovat omassa työnkuvassaan tekemisissä datan kanssa. Tämä johtaa huolellisempaan ja mielekkäämpään työskentelyyn, mikä taas näkyy positiivisesti datan laadussa.

Tavoitteissa on myös mainittu, että kohdeyrityksen tulisi arvioida datalähteitä. Datalähteiden arvioinnissa voidaan hyödyntää esimerkiksi luvussa 2.3.3 esiteltyä IPMAP-työkalua. Datalähteen arvioinnin lisäksi IPMAP:n avulla voidaan mallintaa tietovirtoja, jolloin lopullisen tietotuotteen syntymisen prosessi voidaan ymmärtää paremmin. Tiedon tuotto-prosessin mallintaminen on arvokas osa ongelmien juurisyiden selvittämistä, ja saatujen mallien avulla voidaan tarkastella datan muuttumista/kehittymistä organisaatioyksiköiden ja prosessien yli (Loshin 2011, s. 212).

## 8. PÄÄTELMÄT

Datan määrä monissa yrityksissä on nykyään valtava, ja tämä asettaa erityisiä haasteita datan hallinnalle ja datan laadun varmistamiselle. Datan laatu määrittää, kuinka hyviä yritysten tekemät analyysit ja päätökset voivat olla. Tämän vuoksi datan tehokas hallinta voidaan nähdä erityisen tärkeäksi, sillä sen avulla voidaan osaltaan varmistaa datan korkea laatu. Lisäksi sen avulla voidaan ottaa kantaa kasvavan datan määrän asettamiin haasteisiin. Datan hallinnassa on kuitenkin olennaista ymmärtää, että kyseessä ei ole pelkkä tekninen ongelma, vaan teknologisten näkökulmien lisäksi on huomioitava muun muassa prosessit, käytännöt sekä ihmiset.

Datan laatua voidaan kuvata niin sanottujen datan laadun ulottuvuuksien avulla. Ne ovat laadun osa-alueita, joita voidaan mitata ja ne auttavat ymmärtämään, miksi datan laatua mitataan. Datan laatu voidaankin ymmärtää eri tavoin eri organisaatioissa tai konteksteissa, mutta laadun ulottuvuuksien avulla voidaan edesauttaa yhteistä ymmärrystä datan laadusta. Kaikki ulottuvuudet linkittyvät kuitenkin siihen, mitä dataalta vaaditaan ja mitkä ovat sen käyttäjien tarpeet. Jos data täyttää sille asetetut suorat ja epäsuorat vaatimukset, voidaan sanoa, että datan laatu on lähtökohtaisesti hyvää.

Kirjallisuus tarjoaa useita menetelmiä datan laadun arviointiin. Menetelmiä voidaan erottaa muun muassa sen mukaan, mittaavatko ne datan laatua objektiivisesti vai subjektiivisesti. Datan laadun objektiivinen mittaaminen ei kuitenkaan ole itseisarvo, sillä laatu riippuu vahvasti esimerkiksi sen käyttäjien tarpeista ja kokemuksista. Subjektiiviset mittarit heijastavatkin lähinnä tarpeita ja kokemuksia, kun taas objektiiviset mittarit mahdollistavat datan vertaamisen esimerkiksi johonkin määritettyyn standardiin. Subjektiivisten mittareiden sisällyttäminen arviointiin onkin olennaista datan käyttäjien näkemysten huomioimiseksi.

Datan hallinnoinnilla viitataan siihen, miten dataan ja IT:hen liittyvät vastuut ja prosessit on määritelty sekä miten toimintaa johdetaan. Kyse ei ole siis itse datasta. Datan hallinnointia ei tule myöskään sekoittaa datan hallintaan: hallinnointi on enemmänkin johtamisen viitekehys, kun taas hallinta on jokapäiväistä päätöksentekoa. Datan hallinnoinnin tehokkuutta voidaan arvioida niin sanotuilla kypsyysmalleilla. Kypsyysmalleilla on yhteistä se, että korkeammilla tasoilla korostetaan toiminnan automaatiota ja ongelmien proaktiivista ratkaisemista. Alemmilla tasoilla taas toimintaa ei johdeta eikä se ole suunnitelmallista vaan ennemminkin reaktiivista. Datan hallinnoinnin yhteydessä puhutaan usein myös datan isännöinnistä. Yksinkertaisimmillaan datan isännöinti on datan hallinnoinnissa määriteltujen vastuiden ja prosessien jalkauttamista yrityksen jokapäiväiseen toimintaan.

Prosessilähtöisen datan laadun kehittämisen avulla kehitetään prosesseja, jotka tuottavat ja käsittelevät dataa. Käytännössä tavoitteena on kehittynyt datan hallinta, mikä taas johtaa parempaan datan laatuun. Prosessilähtöinen kehittäminen koostuu viitekehyksistä, onnistumistekijöistä, käytännöistä sekä ohjaavista prosesseista. Viitekehykset auttavat datan hallinnan ongelmien selvittämisessä. Niiden avulla ongelmat voidaan sijoittaa esimerkiksi organisaatiotasojen ja datan elinkaaren vaiheiden rajapintoihin, jolloin ongelmat voidaan paikantaa tarkemmin. Kriittiset onnistumistekijät ovat hankkeiden piirteitä, jotka huomioimalla voidaan merkittävästi vaikuttaa hankkeen onnistumiseen. Parhaat käytännöt taas kuvastavat parhaita toimintatapoja jollain tietyllä alalla tai tietyssä liiketoiminnassa. Prosessimallit puolestaan auttavat huomioimaan hankkeen eri vaiheet sekä mitä missäkin vaiheessa tulisi tehdä. Niiden tehtävä on ohjata hanketta oikeaan suuntaan.

Yrityksille, jotka haluavat kehittää datan hallintaansa, on olemassa paljon erilaisia työkaluja hallinnan kehittämiseen. Tässä työssä käytettyjä työkaluja ovat muun muassa datan laadun arvioinnin menetelmät, datan hallinnoinnin kypsyysmallit ja viitekehykset sekä POSMAD-viitekehys datan hallinnan ongelmien löytämiseen. Kun yritys lähtee suunnittelemaan datan hallinnan kehittämisen hanketta, voi se käyttää lisäksi erilaisia prosessimalleja suunnitellakseen hankkeen kulun. Arviointimenetelmiä ja viitekehyksiä voidaan taas käyttää prosessin eri vaiheissa. Myös vahvasti vakiintuneet viitekehykset, kuten COBIT, tarjoavat yrityksille apuvälineitä datan hallinnan kehittämiseen. Näitä voidaan käyttää soveltuvien osien vastaamaan yrityksen tarpeita.

Tässä työssä datan hallinnan kehittämis ehdotukset luotiin seuraavasti. Työn tekijä tutustui kohdeyrityksen dataan ja sen hallintaan liittyviin käytäntöihin. Tämän jälkeen tunnistettiin kriittiset alueet ja rajattiin tarkasteltava data. Datan laadun arviointi varten valittiin sopivat laadun ulottuvuudet, määriteltiin haastattelukysymykset ja -tekniikat sekä valittiin haastateltavat henkilöt. Tämän jälkeen tiedot kerättiin ja tulokset analysoitiin. Samoilta haastatelluilta kerättiin näkemyksiä myös datan hallinnan ongelmista. Tulokset lajiteltiin POSMAD-viitekehyksen matriisiin, minkä jälkeen kohdeyritykselle määriteltiin datan hallinnan kypsyys nyky- ja tavoitetaso käyttäen erilaisia datan hallinnan kypsyysmalleja. Lopulliset kehitysehdotukset luotiin kirjallisuuden, havaittujen ongelmien sekä määritellyn tavoitetason avulla. Datan laadun arvioinnin tuloksia voidaan käyttää esimerkiksi referenssitasona vastaisuudessa suoritettavalle datan laadun arvioinnille. Näin voidaan selvittää, ovatko datan hallinnan ja laadun kehittämisen hankkeen hyödyt realisoituneet parantuneena datan laaduna.

Vakiintuneista datan laadun arviointimenetelmistä AIMQ on kevyin toteuttaa. Siinä on kuitenkin selkeitä rajoitteita, esimerkiksi arvioinnin puhdas subjektiivisuus sekä benchmarking-analyysin vaikea toteuttaminen. Useat muut arviointi- ja mittausmenetelmät vaativat mittareiden integroimisen joko prosesseihin tai tietojärjestelmiin. Näitä voi olla vaikea toteuttaa pienillä resursseilla ja lyhyellä aikavälillä. Ne kuitenkin mahdollistavat datan laadun jatkuvan seurannan ja mittaamisen.



Datan hallinnoinnin ehdotetut viitekehykset soveltuvat hyvin isoille yhtiöille, joilla on suuret resurssit ja suuri keskijohto. Niiden soveltaminen suomalaisiin pieniin ja keskikokoisiin yrityksiin ei ole aivan ongelmatonta. Esimerkiksi datan hallinnoinnin ohjelmaa suunnitellessaan yritys voi joutua muokkaamaan viitekehyksiä niin, että ne sopivat paremmin yrityksen toimintaan. Viitekehykset sellaisinaan voivat olla liian raskaita ja joustamattomia pienissä yrityksissä. Esimerkiksi pienet yritykset eivät välttämättä vaadi erillistä datan hallinnoinnin komiteaa. Tosin jokin muu strateginen ryhmä voi hoitaa datan hallinnoinnin komitealle kuuluvia tehtäviä muiden tehtävien ohella.

Kirjallisuus tarjoaa myös paljon keinoja datan laadun kehittämiseen, mutta vähemmän on tutkittu juuri datan hallinnan kehittämistä. Tämä tarkoittaa sitä, että monet yritykset joutuvat itse luomaan datan hallinnan kehittämisen suunnitelmat yhdistämällä esimerkiksi datan hallinnoinnin ja datan laadun kehittämisen teoriaa. Kuitenkin esimerkiksi tässä työssä esitellyt prosessilähtöiset keinot laadun kehittämiseen ovat melkein suoraan datan hallinnan kehittämisen keinoja.

Työssä saatujen tulosten mukaan kohdeyrityksen datan laatu on pääosin hyvä. Vain helpokäyttöisyyden kohdalla laatu koetaan selkeästi heikoksi. Toisaalta joidenkin datan laadun ulottuvuuksien toteutuminen nähdään erittäin hyväksi (esimerkiksi datan maine). Datan hallinnassa koetaan kuitenkin olevan enemmän ongelmia kuin datan laadussa. Datan hallintaan liittyvät ongelmat jakaantuvat melko tasaisesti POSMAD-matriisiin. Tämä yhdessä datan hallinnan nykytason määrittelyn kanssa vahvistaa päätelmää, että kohdeyrityksen datan hallinta on alkeellista eikä spesifisten kehitysehdotuksien antaminen ole taroituksenmukaista. Jos taas ongelmat sijoittuisivat jollekin sarakkeelle, riville tai tiettyihin soluihin, on helpompi luoda tarkempia ehdotuksia ongelmien korjaamiseksi.

Tutkimus onnistuu vastaamaan asetettuihin tutkimuskysymyksiin. Päättötutkimuskysymykseen vastattiin teorialukujen avulla ja siihen palattiin myös päätelmissä. Työn alakysymyksiin vastattiin puolestaan empirialukujen avulla. Kohdeyrityksen kannalta työn tärkeimmät tulokset ovat datan hallinnan kehitysehdotukset. Nämä on pyritty esittämään mahdollisimman yksityiskohtaisesti ja motivoivasti. On kuitenkin syytä korostaa, että tämän työn tulokset ovat vain kehitysehdotuksia. Vastuu niiden jalkauttamisesta on kohdeyrityksellä, jonka tuleekin pohtia, miten tässä työssä esitetyt kehitysehdotukset implementoidaan omassa organisaatiossa tai miten niitä jatkokehitetään. Datan hallinnan kehittämiseen ei ole nimittäin yhtä valmista ratkaisua, joka sopii jokaiselle organisaatiolle. Tämän vuoksi yritysten tuleekin käyttää olemassa olevia työkaluja ja parhaita käytäntöjä luodakseen sellaisen datan hallinnan kehittämisen ohjelman, joka parhaiten sopii omaan organisaatioon.

Tämän työn tulokset voivat olla kohdeyritykselle hyvinkin merkittäviä. Jos datan hallinta koetaan liiketoiminnallisesti tärkeäksi toiminnaksi, on näistä tuloksista suuri hyöty datan hallinnan kehittämisessä. Ne auttavat yritystä luomaan suunnitelman datan hallinnan ke-

hittämiseksi sekä antavat ehdotuksia kehitystoimenpiteiksi. Lisäksi työssä on esitelty paljon hyviä datan hallintaan liittyviä käytäntöjä, menetelmiä ja viitekehyksiä. Tuloksia voidaan myös jollain tasolla yleistää kohdeyrityksen ulkopuolella erityisesti pieni- ja keskikokoisiin yrityksiin. Pienemmät yritykset keskittyvät usein ydinliiketoimintaan, jolloin tukitoimintojen, esimerkiksi datan hallinnan, saama huomio voi olla vähäistä. Jos yrityksen ydinliiketoiminta kuitenkin vaatii datan hallintaa toimiakseen tehokkaasti, ovat tämän työn tulokset hyvinkin relevantteja. Lisäksi yritykset, jotka suunnittelevat datan hallintaa tai pohtivat sen kehittämistä, voivat kokea työn tulokset hyödyllisiksi. Työn kontribuutio datan hallinnan alalle on sen sijaan melko pieni, sillä työssä ei ole varsinaisesti tuotettu uusia näkemyksiä tai synteesejä liittyen datan hallintaan. Työssä on kuitenkin käytetty kirjallisuudessa esiteltyjä menetelmiä ja arvioitu niiden toimivuutta. Lisäksi työssä on yhdistetty datan hallinnan eri malleja tulosten analysoinnissa ja kehitysehdotusten luomisessa. Uusien näkemysten tuottaminen ei kuitenkaan ollut tämän työn tarkoitus, vaan työ on pikemminkin pragmaattinen lisä datan hallinnan alalle.

Työssä saadut tulokset ovat keskenään erittäin johdonmukaisia, mikä lisää niiden uskottavuutta ja vakuuttavuutta. Eri roolien näkemykset ovat hyvin samanlaisia, ja samoja asioita nousi esille useissa haastatteluissa. Työssä haastateltiin myös datan käyttäjiä kohdeyrityksen ulkopuolelta. Näin oli mahdollista kerätä näkemyksiä, joita ei ehkä muuten olisi tuotu esille. Tulosten analysoinnissa on käytetty useita kirjallisuuden tarjoamia menetelmiä. Esimerkiksi datan hallinnan nykytason määrittelyssä on käytetty neljää eri kypsyysmallia. Lisäksi datan laadun tulosten keräämiseen on käytetty kirjallisuudessa esiteltyä arviointimenetelmää.

Työn onnistumista voidaan arvioida myös reliabiliteetin ja validiteetin suhteen. Yin (2009, s. 40) esittää neljä testiä, joiden avulla voidaan arvioida tapaustutkimusten onnistumista. Rakenteellinen validiteetti on ilmiön tutkimiseen sopivien menetelmien tunnistamista. Sisäinen validiteetti taas liittyy syy-seuraussuhteiden luomiseen. Ulkoinen validiteetti kuvaa, miten tutkimuksen tulosten yleistettävyyden on määriteltävä. Reliabiliteetti liittyy siihen, miten hyvin työ on toistettavissa saaden samat tulokset. (Yin 2009, s. 40.) Nämä termit, niihin liittyvät tapaustutkimusten taktiikat sekä tässä työssä käytetyt menetelmät on esitetty taulukossa 20.

Tässä työssä käytetyt menetelmät työn rakenteellisen validiteetin takaamiseksi ovat haastattelut ja kirjallisuus. Haastattelut nauhoitettiin, niiden aikana tehtiin muistiinpanoja ja lisäksi haastattelut litteroitiin. Useat osallistujat kohdeyrityksessä myös arvioivat työn tulosten oikeellisuutta. Sisäiseen validiteettiin liittyen työn haastatteluissa löydettiin toistuvia kaavoja tai asioita. Joitain syy-seuraussuhteita pystyttiin myös muodostamaan. Kilpaileviin selityksiin ei voitu ottaa kantaa, sillä näitä ei löydetty. Logiikkamalleja ei myöskään käytetty, koska ne vaativat ilmiön pitkän aikavälin tarkastelua. Ulkoinen validiteetti pyrittiin varmistamaan käyttämällä olemassa olevaa kirjallisuutta kuvaamaan nykyistä tilannetta sekä luodessa kehitysehdotuksia. Työn reliabiliteettiin liittyen työlle määriteltiin

selkeä prosessi ja kaikissa haastatteluissa käytettiin samaa haastattelupohjaa. Lisäksi kaikki työssä käytetty materiaali on talletettu myöhempää tarkastelua varten.

**Taulukko 20.** Tapaustutkimuksen testit ja työn arviointi (mukaillen Yin 2009, s. 41)

Testi	Tapaustutkimuksen taktiikka	Tässä työssä käytetyt menetelmät
Rakenteellinen validiteetti	Käytä useita todisteiden lähteitä	Haastattelut ja kirjallisuus
	Luo todisteiden ketju	Haastattelut nauhoitettiin ja niissä tehtiin muistiinpanoja. Haastattelut litteroitiin
	Anna tärkeiden tiedonantajien arvioida tutkimuksen raportin luonnos	Useat osallistujat kohdeyrityksessä arvioivat työtä ja sen tuloksia
Sisäinen validiteetti	Yhdistä toistuvia malleja/kaavoja	Toistuvia kaavoja/asioita löydettiin haastatteluissa
	Luo selityksiä	Joitain syy-seuraussuhteita löydetty
	Ota kantaa kilpaileviin selityksiin	Ei käytetty – ei löydetty kilpailevia selityksiä
	Käytä logiikkamalleja	Ei käytetty – vaatii ilmiön pitkän aikavälin tarkastelua
Ulkoinen validiteetti	Yhden tapauksen tutkimuksissa luo teoria	Olemassa olevaa teoriaa käytetty kuvaamaan nykyistä tilannetta sekä luodessa kehitysehdotuksia
	Luo vastaavia teorioita useamman tapauksen tutkimuksissa	Ei käytetty – kyseessä yhden tapauksen tutkimus
Reliabiliteetti	Käytä tapaustutkimuksen protokollaa	Työlle määritelty selkeä prosessi, haastatteluissa käytetty samaa pohjaa
	Kehitä tapaustutkimuksen tietokanta	Haastattelujen äänitteet, puhtaaksi kirjoitetut tekstit, muistiinpanot, kirjallisuus sekä muu käytetty materiaali talletettu

Työn rajoitteet liittyvät pitkälti datan laadun arvioinnin subjektiivisuuteen ja laadullisuuteen. Objektiivisten mittareiden käyttäminen laadun arviointiin olisi ollut tämän työn puitteissa hankalaa, sillä objektiiviset mittarit usein integroidaan dataan liittyviin prosesseihin tai tietojärjestelmiin. Kvantitatiivisella laadun arvioinnilla ei oltaisi taas saatu riittävää tietoa laadun ulottuvuuksiin liittyvistä ongelmista. Datan laadun arvioinnissa tehty valinnat kuitenkin heikentävät tulosten vertailukelpoisuutta tulevaisuudessa tehtäviin arviointeihin.

Jatkossa datan hallintaan liittyvien tutkimusten tulisi keskittyä erityisesti datan hallinnan kehittämiseen pienissä ja keskikokoisissa yrityksissä esimerkiksi case-yritysten avulla. Näin vakiintuneita malleja ja työkaluja voidaan testata pienemmissä yrityksissä. Tämä auttaa selvittämään, miten hyvin mallit toimivat pienissä organisaatioissa sekä miten malleja tulisi soveltaa pienten yritysten kontekstissa. Yleensäkin datan hallintaan liittyviä pragmaattisia tutkimuksia ja teoksia on saatavilla hyvin vähän.

## LÄHTEET

Adler, S. Six Steps to Data Governance Success. CIO. 31.5.2007 [viitattu 29.10.2015].  
Saataavissa: <http://www.cio.com/article/2438861/enterprise-architecture/six-steps-to-data-governance-success.html>.

Ballou, D., Wang, R., Pazer, H. & Kumar.Tayi, G. 1998. Modeling Information Manufacturing Systems to Determine Information Product Quality. *Management Science* 44, 4, pp. 462-484.

Batini, C., Cabitza, F., Cappiello, C. & Francalanci, C. 2008. A Comprehensive Data Quality Methodology for Web and Structured Data. *International Journal of Innovative Computing and Applications* 1, 3, pp. 205-218.

Batini, C., Cappiello, C., Francalanci, C. & Maurino, A. 2009. Methodologies for Data Quality Assessment and Improvement. *ACM Computing Surveys* 41, 3, 52 p.

Beath, C., Becerra-Fernandez, I., Ross, J. & Short, J. 2012. Finding Value in the Information Explosion. *MIT Sloan Management Review* 53, 4, pp. 18-20.

Bobrowski, M., Marré, M. & Yankelevic, D. 1999. Measuring Data Quality. Buenos Aires, Argentina, Universidad de Buenos Aires. Report 99-002. 11 p.

Brennan, R. From Chaos to Nirvana in Data Governance Evolution. *TabbFORUM*. 21.2.2012 [viitattu 17.9.2015]. Saataavissa: <http://tabbforum.com/opinions/from-chaos-to-nirvana-in-data-governance-evolution>.

BusinessDictionary. Best practice definition. WebFinance, Inc. [viitattu 30.11.2015a]. Saataavissa: <http://www.businessdictionary.com/definition/best-practice.html>.

BusinessDictionary. Framework definition. WebFinance, Inc. [viitattu 11.12.2015b]. Saataavissa: <http://www.businessdictionary.com/definition/framework.html>.

Caballero, I., Verbo, E., Calero, C. & Piattini, M. 2007. A Data Quality Measurement Information Model Based on ISO/IEC 15939. *The 12th International Conference on Information Quality*, 9.-11.11.2007. Massachusetts Institute of Technology. 16 p.

Cambridge Dictionaries Online. Best practice definition. Cambridge University Press [viitattu 30.11.2015]. Saataavissa: <http://dictionary.cambridge.org/dictionary/english/best-practice>.

Cheong, L.K. & Chang, V. 2007. The Need for Data Governance: A Case Study. *ACIS 2007 Proceedings*, 5.-8.12.2007. AIS Electronic Library. pp. 999-1008.

Cohen, R. What's in a Name? Data Governance Roles, Responsibilities and Results Factors. Information Management Online. 8.6.2006 [viitattu 17.9.2015]. Saatavissa: <http://www.information-management.com/news/columns/-1057220-1.html>.

Davidson, B., Lee, Y.W. & Wang, R. 2004. Developing data production maps: meeting patient discharge data submission requirements. International Journal of Healthcare Technology and Management 6, 2, pp. 223-240.

DGIQ IQ-Definition. Deutsche Gesellschaft für Informations- und Datenqualität. 2007 [viitattu 28.10.2015]. Saatavissa: [http://www.competence-site.de/content/uploads/1f/17/IQ-Def\\_GIQMC2-Version-AP.pdf](http://www.competence-site.de/content/uploads/1f/17/IQ-Def_GIQMC2-Version-AP.pdf).

Friedman, T. 2006. Key Issues for Data Management and Integration. Gartner Research 30.3.2006, 6 p.

Helfert, M., Foley, O., Ge, M. & Cappiello, C. 2009. Analysing the Effect of Security on Information Quality Dimensions. ECIS 2009 Proceedings, 8.-10.6.2009. AIS Electronic Library. 13 p.

IBM 2007. The IBM Data Governance Council Maturity Model: Building a roadmap for effective data governance. Somers, NY, IBM Software Group. 16 p.

ITGI 2007. COBIT 4.1. Rolling Meadows, IL, USA, IT Governance Institute. 196 p.

Kahn, B.K., Strong, D.M. & Wang, R.Y. 2002. Information Quality Benchmarks: Product and Service Performance. Communications of the ACM 45, 4, pp. 184-192.

Karel, R. Data Governance Framework Walkthrough: Vision and Business Case. Informatica. 2.1.2014a [viitattu 2.11.2015]. Saatavissa: <http://blogs.informatica.com/2014/01/02/data-governance-framework-walkthrough-vision-and-business-case/#fbid=DHq5ps14V7I>.

Karel, R. The Process Stages of Data Governance. Informatica. 2.1.2014b [viitattu 2.11.2015]. Saatavissa: <http://blogs.informatica.com/2014/01/02/the-process-stages-of-data-governance/#fbid=DHq5ps14V7I>.

Khatri, V. & Brown, C.V. 2010. Designing Data Governance. Communications of the ACM 53, 1, pp. 148-152.

Kovac, R., Lee, Y.W. & Pipino, L.L. 1997. Total Data Quality Management: The Case of IRI. The 1997 Conference on Information Quality, 24.-26.10.1997. Massachusetts Institute of Technology. pp. 63-79.

Kovac, R. & Weickert, C. 2002. Starting with Quality: Using TDQM in a Start-Up Organization. Proceedings of the Seventh International Conference on Information Quality, 8.-10.11.2002. Massachusetts Institute of Technology. pp. 69-78.

Laatikainen, T. & Niemi, E. 2012. Data Quality johtavissa suomalaisyrityksissä: nykytila ja tarvittavat rakenteet. DAMA Finland ry. White paper. 9 p.

Ladley, J. 2012. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program. 1st ed. Waltham, MA, Elsevier. 236 p.

Lee, Y.W., Strong, D.M., Kahn, B.K. & Wang, R.Y. 2002. AIMQ: a methodology for information quality assessment. *Information & Management* 40, pp. 133-146.

Leidecker, J.K. & Bruno, A.V. 1984. Identifying and using critical success factors. *Long Range Planning* 17, 1, pp. 23-32.

Loshin, D. 2011. The Practitioner's Guide to Data Quality Improvement. 1st ed. Burlington, MA, Elsevier. 398 p.

Lucas, A. 2010. Corporate Data Quality Management: From Theory to Practice. 5th Iberian Conference on Information Systems And Technologies, 16.-19.6.2010. IEEE. 7 p.

Malange, S.N., Ngassam, E.K., Ojo, S.O. & Osunmakinde, I.O. 2015. Methodology for Improving Data Quality Management in South African Government Departments. IST-Africa 2015, IIMC. 8 p.

Marinos, G. We're Not Doing What? The Top 10 Corporate Oversights in Data Governance. *Information Management Online*. 1.9.2004 [viitattu 17.9.2015]. Saatavissa: <http://www.information-management.com/issues/20040901/1009205-1.html>.

McGilvray, D. 2008. Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information. 1st ed. Burlington, MA, Elsevier. 334 p.

Moges, H.T. 2014. A Contextual Data Quality Analysis for Credit Risk Management in Financial Institutions. Doctor of Applied Economics. Katholieke Universiteit Leuven. 177 p.

Newman, D. & Logan, D. 2006. Governance Is an Essential Building Block for Enterprise Information Management. Gartner Research 18.5.2006, 9 p.

Niemi, E. & Kontra, K. 2012. Data Governance johtavissa suomalaisyrityksissä: nykytila ja yhdeksän suositusta. DAMA Finland ry. White paper. 6 p.

Page, J. 2011. How to Launch a Data Governance Initiative. *Business Intelligence Journal* 16, 2, pp. 17-25.

- Panian, Z. 2010. Some Practical Experiences in Data Governance. *World Academy of Science, Engineering and Technology* 62, pp. 939-946.
- Pipino, L.L., Lee, Y.W. & Wang, R.Y. 2002. Data Quality Assessment. *Communications of the ACM* 45, 4, pp. 211-218.
- Plotkin, D. 2014. *Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance*. 1st ed. Waltham, MA, Elsevier. 224 p.
- Power, D. 2011. How to Start a Data Governance Program. *Information Management* Mar/Apr, pp. 31-32.
- Rickards, R.C. & Ritsert, R. 2012. Data Governance Challenges Facing Controllers. *International Journal of Business, Accounting, and Finance* 6, 1, pp. 25-42.
- Rouse, M. Unstructured data definition. TechTarget [viitattu 27.11.2015]. Saatavissa: <http://searchbusinessanalytics.techtarget.com/definition/unstructured-data>.
- Saaranen-Kauppinen, A. & Puusniekka, A. KvaliMOTV - Menetelmäopetuksen tietovaranto. Yhteiskuntatieteellinen tietoaarkisto. Tampere. 2006 [viitattu 17.9.2015]. Saatavissa: <http://www.fsd.uta.fi/menetelmaopetus/>.
- Saunders, M., Lewis, P. & Thornhill, A. 2009. *Research methods for business students*. 5th ed. Essex, UK, Pearson Education Ltd. 614 p.
- Sebastian-Coleman, L. 2013. *Measuring Data Quality for Ongoing Improvement*. Waltham, MA, USA, Elsevier, Inc. 370 p.
- Shankaranarayanan, G. & Cai, Y. 2006. Supporting data quality management in decision-making. *Decision Support Systems* 42, pp. 302-317.
- Shankaranarayanan, G., Wang, R.Y. & Ziad, M. 2000. IP-MAP: Representing the Manufacture of an Information Product. *Proceedings of the 2000 Conference on Information Quality*, 20.-22.10.2000. Massachusetts Institute of Technology. pp. 1-16.
- Shankaranarayanan, G., Ziad, M. & Wang, R.Y. 2003. Managing Data Quality in Dynamic Decision Environments: An Information Product Approach. *Journal of Database Management* 14, 4, pp. 14-32.
- Storey, V.C., Dewan, R.M. & Freimer, M. 2012. Data quality: Setting organizational policies. *Decision Support Systems* 54, pp. 434-442.
- Tee, S.W., Bowen, P.L., Doyle, P. & Rohde, F.H. 2007. Factors influencing organizations to improve data quality in their information systems. *Accounting and Finance* 47, pp. 335-355.

- Thomas, G. 2006. Alpha Males and Data Disasters: The Case for Data Governance. USA, Brass Cannon Press. 222 p.
- Todd, G. 2008. Data Governance: The Enabler of High Performance. DM Review May, pp. 30-31.
- Tuomi, J. & Sarajärvi, A. 2002. Laadullinen tutkimus ja sisällönanalyysi. Helsinki, Tammi.
- Waddington, D. 2010. Data Governance, MDM and Data Quality. Information Management Sept/Oct, pp. 14-16.
- Wang, K.Q., Tong, S.R., Roucoules, L. & Eynard, B. 2008. Analysis of Data Quality and Information Quality Problems in Digital Manufacturing. Proceedings of the 2008 IEEE ICMIT, IEEE. pp. 439-443.
- Wang, R.Y. & Strong, D.M. 1996. Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems 12, 4, pp. 5-33.
- Wang, R.Y. 1998. A Product Perspective on Total Data Quality Management. Communications of the ACM 41, 2, pp. 58-65.
- Weber, K., Otto, B. & Österle, H. 2009. One Size Does Not Fit All - A Contingency Approach to Data Governance. ACM Journal of Data and Information Quality 1, 1, 27 p.
- Wende, K. 2007. A Model for Data Governance - Organising Accountabilities for Data Quality Management. 18th Australasian Conference on Information Systems, 5.-7.12.2007. AIS Electronic Library. pp. 417-425.
- Xu, H. 2003. Critical Success Factors for Accounting Information Systems Data Quality. Doctor of Philosophy. University of Southern Queensland. 233 p.
- Yin, R.K. 2009. Case Study Research: Design and Methods. 4th ed. Thousand Oaks, CA, SAGE Publications. 219 p.



## LIITE A: HAASTATTELUPOHJA

### Datan laadun arviointi

1. *Saatavuus*
  - a. Onko data helposti saatavissa?
  - b. Onko data nopeasti saatavissa?
  - c. Onko data helposti palautettavissa?
2. *Sopiva määrä*
  - a. Onko datan määrä sopiva tarpeisiisi?
3. *Luotettavuus*
  - a. Onko data luotettavaa?
4. *Täydellisyys*
  - a. Sisältääkö data kaikki tarvittavat arvot?
  - b. Onko datan syvyys/yksityiskohtaisuus riittävä tehtäviisi nähden?
5. *Tiivis esitystapa*
  - a. Onko data muotoiltu tiiviisti?
  - b. Onko data esitetty tiiviisti?
  - c. Onko tarvittavan datan ohella esitetty tarpeetonta dataa?
6. *Johdonmukainen esitystapa*
  - a. Onko data esitetty johdonmukaisesti samassa muodossa?
7. *Helppokäyttöisyys*
  - a. Onko dataa helppo muokata vastaamaan tarpeitasi?
  - b. Onko dataa helppo yhdistää muun datan kanssa?
8. *Oikeellisuus/tarkkuus*
  - a. Onko data virheetöntä?
  - b. Onko data riittävän tarkkaa?
9. *Tulkittavuus*
  - a. Ovatko datan mittausyksiköt selkeät?
  - b. Onko data helposti tulkittavissa?
10. *Objektiivisuus*
  - a. Onko data kerätty objektiivisesti?
  - b. Perustuuko data faktoihin?
  - c. Esittääkö data puolueettoman näkökulman?
11. *Relevanssi/hyödyllisyys*
  - a. Onko data hyödyllistä työllesi?
  - b. Soveltuuko data työhösi?
12. *Maine*
  - a. Onko datalla hyvä maine?
  - b. Tuleeko data hyvistä lähteistä?
13. *Tietoturvallisuus*
  - a. Suojataanko data luvattomalta pääsylvä?

- b. Onko pääsyä dataan rajoitettu sopivasti?
- c. Pääsevätkö dataan käsiksi vain henkilöt, joiden kuuluu nähdä se?

*14. Oikea-aikaisuus*

- a. Onko data riittävän tuoretta työhösi?

*15. Ymmärrettävyys*

- a. Onko datan merkitys helppo ymmärtää?

**Datan hallinnan arviointi**

- 1. Mitä ongelmia datan hallinnassa mielestäsi on?*
- 2. Mitä haasteita datan hallintaan mielestäsi liittyy?*
- 3. Miten datan hallintaa tulisi mielestäsi kehittää?*

## LIITE B: AIMQ-KYSELYLOMAKE

Kaikki kohteet ovat Lee et al. (2002, s. 143-144) mukaan. Kohteet merkitty (K) ovat käänteisiä väitteitä.

### Saatavuus

Tämä data on helposti palautettavissa.

Tämä data on helposti saatavissa.

Tämä data on helposti hankittavissa.

Tämä data on tarvittaessa nopeasti saatavissa.

### Sopiva määrä

Tämän datan määrä on sopiva tarpeisiimme.

Datan määrä ei vastaa tarpeitamme. (K)

Datan määrä ei ole riittävä tarpeisiimme nähden. (K)

Datan määrä ei ole liian suuri eikä liian pieni.

### Uskottavuus

Tämä data on uskottavaa.

Tämän datan uskottavuus on kyseenalaista.

Tämä data on luotettavaa.

### Täydellisyys

Tämä data sisältää kaikki tarvittavat arvot.

Tämä data on epätäydellistä. (K)

Tämä data on täydellistä.

Tämä data on riittävän täydellistä tarpeisiimme.

Tämä data täyttää tehtävienne tarpeet.

Tämän datan syvyys on riittävä tehtäviimme nähden.

### Tiivis esitystapa

Tämä data on muotoiltu tiiviisti.

Tämä data on esitetty tiiviisti.

Tämä data on esitetty tiiviissä muodossa.

Tämän datan esitys on tiivis.

### Johdonmukainen esitystapa

Tämä data on johdonmukaisesti esitetty samassa muodossa.

Tätä dataa ei ole esitetty johdonmukaisesti. (K)

Tämä data on esitetty johdonmukaisesti.

Tämä data on esitetty johdonmukaisessa muodossa.

### Helppokäyttöisyys

Tätä dataa on helppo muokata vastaamaan tarpeitamme.

Tätä dataa on helppo yhdistää.

Tätä dataa on vaikea muokata vastaamaan tarpeitamme. (K)

Tätä dataa on vaikea yhdistää. (K)

Tätä dataa on helppo yhdistää muun datan kanssa.

### Oikeellisuus

Tämä data on virheetöntä.

Tämä data on virheellistä. (K)

Tämä data on tarkkaa.

Tämä data on luotettavaa.

#### Tulkittavuus

Tämän datan merkitystä on helppo tulkita.  
Tätä dataa on vaikea tulkita. (K)  
Tämä data on helposti tulkittavissa.  
Tämän datan mittaussyksiköt ovat selkeät.

#### Objektiivisuus

Tämä data on kerätty objektiivisesti.  
Tämä data perustuu faktoihin.  
Tämä data on objektiivista.  
Tämä data esittää puolueettoman näkökulman.

#### Relevanssi

Tämä data on hyödyllistä työllemme.  
Tämä data on relevanttia työllemme.  
Tämä data on sopivaa työllemme.  
Tämä data soveltuu työhömmme.

#### Maine

Tämän datan laadun maine on huono. (K)  
Tällä datalla on hyvä maine.  
Tämän datan laadulla on hyvä maine.  
Tämä data tulee hyvistä lähteistä.

#### Tietoturvallisuus

Tätä dataa suojataan luvattomalta pääsylvä.  
Tätä dataa ei ole suojattu riittäväällä tietoturvalla. (K)  
Pääsyä tähän dataan on rajoitettu sopivasti.  
Tähän dataan pääsee käsiksi vain henkilöt, joiden kuuluu nähdä se.

#### Oikea-aikaisuus

Tämä data on riittävän ajankoh-  
taista työllemme.  
Tämä data ei ole riittävän ajan-  
kohtaista. (K)  
Tämä data on riittävän oikea-  
aikaista.  
Tämä data on riittävän tuoretta  
työllemme.

#### Ymmärrettävyys

Tätä dataa on helppo ymmärtää.  
Tämän datan merkitystä on vai-  
kea ymmärtää. (K)  
Tämän datan merkitys on  
helppo ymmärtää.

## LIITE C: POSMAD-VIITEKEHYS

	Suunnittelu	Kerääminen	Tallettaminen ja jakaminen
<b>Data (mitä)</b>	<ul style="list-style-type: none"> <li>-Mitkä ovat liiketoimintatavoitteet?</li> <li>-Mikä data tukee liiketoimintatarpeita?</li> <li>-Mitkä ovat liiketoimintasäännöt?</li> <li>-Mitkä ovat datastandardit?</li> </ul>	<ul style="list-style-type: none"> <li>-Mitä dataa kerätään?</li> <li>-Mitä dataa syötetään järjestelmään - yksittäiset dataelementit tai uudet arkistot?</li> </ul>	<ul style="list-style-type: none"> <li>-Mitä dataa talletetaan?</li> <li>-Mitä dataa jaetaan?</li> <li>-Mikä on avaindataa, joka varmuuskopioidaan nopean saatavuuden varmistamiseksi?</li> </ul>
<b>Prosessit (miten)</b>	<ul style="list-style-type: none"> <li>-Mitkä ovat korkean tason prosessit?</li> <li>-Millainen on koulutuksen ja viestinnän strategia?</li> </ul>	<ul style="list-style-type: none"> <li>-Miten data kerätään eri lähteistä (sisäiset ja ulkoiset)?</li> <li>-Miten data syötetään järjestelmään?</li> <li>-Mitkä tekijät vaikuttavat uusien arkistojen luomiseen?</li> </ul>	<ul style="list-style-type: none"> <li>-Millainen on datan tallettamisen prosessi?</li> <li>-Millainen on datan jakamisen prosessi?</li> </ul>
<b>Ihmiset ja organisaatiot (mitä)</b>	<ul style="list-style-type: none"> <li>-Kuka tunnistaa liiketoimintatavoitteet ja osoittaa prioriteetit ja resurssit?</li> <li>-Kuka kehittää prosessit, liiketoimintasäännöt ja standardit?</li> <li>-Kuka johtaa tähän vaiheeseen kuuluvia henkilöitä?</li> </ul>	<ul style="list-style-type: none"> <li>-Kuka kerää dataa lähteistä?</li> <li>-Kuka syöttää uutta dataa ja luo uusia arkistoja järjestelmään?</li> <li>-Kuka johtaa tähän vaiheeseen kuuluvia henkilöitä?</li> </ul>	<ul style="list-style-type: none"> <li>-Kuka tukee/ylläpitää tallettamiseen liittyvää teknologiaa?</li> <li>-Kuka tukee jakamiseen liittyvää teknologiaa?</li> <li>-Kuka johtaa tähän vaiheeseen kuuluvia henkilöitä?</li> </ul>
<b>Teknologia (miten)</b>	<ul style="list-style-type: none"> <li>-Millainen on korkean tason arkkitehtuuri ja liiketoimintaa tukevat teknologiat?</li> </ul>	<ul style="list-style-type: none"> <li>-Miten sovellusta käytetään luomaan uutta tietoa ja uusia arkistoja järjestelmään?</li> </ul>	<ul style="list-style-type: none"> <li>-Millaista teknologiaa datan tallettamiseen käytetään?</li> <li>-Millaista teknologiaa datan jakamiseen käytetään?</li> </ul>

	Ylläpito	Käyttö	Hävittäminen
<b>Data (mitä)</b>	<ul style="list-style-type: none"> <li>-Mitä dataa päivitetään ja muutetaan järjestelmässä?</li> <li>-Mitä dataa muutetaan ennen migraatiota, integraatiota tai jakamista?</li> <li>-Mitä dataa aggregoidaan mittarien ja raportoinnin tukemiseksi?</li> </ul>	<ul style="list-style-type: none"> <li>-Mitä dataa liiketoiminta tarvitsee tukeakseen liiketapahtumia, mittareita, standardien noudattamista vaatimuksia, päätöksentekoa, automatisoituja prosesseja ja muita tavoitteita?</li> <li>-Mitä dataa on saatavilla liiketoiminnalle?</li> </ul>	<ul style="list-style-type: none"> <li>-Mitä dataa tulee arkistoida?</li> <li>-Mitä dataa tulee poistaa?</li> </ul>
<b>Prosessit (miten)</b>	<ul style="list-style-type: none"> <li>-Miten dataa päivitetään?</li> <li>-Miten dataa seurataan muutosten huomaimiseksi?</li> <li>-Miten standardeja ylläpidetään?</li> <li>-Miten datan muutosta hallitaan ja seuraukset arvioidaan?</li> <li>-Mitkä tekijät vaikuttavat ylläpitoon?</li> </ul>	<ul style="list-style-type: none"> <li>-Miten dataa käytetään?</li> <li>-Miten dataan pääsee käsiksi?</li> <li>-Miten datan tietoturvasuus on varmistettu?</li> <li>-Miten data saadaan saatavaksi sen käyttäjille?</li> <li>-Mitkä tekijät vaikuttavat datan käyttöön?</li> </ul>	<ul style="list-style-type: none"> <li>-Miten dataa arkistoidaan?</li> <li>-Miten dataa poistetaan?</li> <li>-Miten arkistosijainteja ja prosesseja johdetaan?</li> <li>-Mitkä ovat arkistointiin ja hävittämiseen vaikuttavat tekijät?</li> </ul>
<b>Ihmiset ja organisaatiot (mitä)</b>	<ul style="list-style-type: none"> <li>-Kuka päättää mitä päivitetään?</li> <li>-Kuka tekee varsinaiset muutokset järjestelmässä?</li> <li>-Kuka on vastuussa datan laadusta?</li> <li>-Keiden tulee tietää muutoksista?</li> <li>-Kuka johtaa tähän vaiheeseen kuuluvia henkilöitä?</li> </ul>	<ul style="list-style-type: none"> <li>-Kuka käsittelee dataa suoraan?</li> <li>-Kuka käyttää dataa?</li> <li>-Kuka johtaa tähän vaiheeseen kuuluvia henkilöitä?</li> </ul>	<ul style="list-style-type: none"> <li>-Kuka asettaa datan säilyttämisen käytännöt/linjaukset?</li> <li>-Kuka arkistoi datan?</li> <li>-Kuka poistaa datan?</li> <li>-Keille arkistoinnista / poistamisesta tulee ilmoittaa?</li> <li>-Kuka johtaa tähän vaiheeseen kuuluvia henkilöitä?</li> </ul>
<b>Teknologia (miten)</b>	<ul style="list-style-type: none"> <li>-Miten dataa ylläpidetään ja päivitetään järjestelmässä?</li> </ul>	<ul style="list-style-type: none"> <li>-Miten dataan pääsee käsiksi eri liiketoimintatarpeiden tyydyttämiseksi?</li> <li>-Miten liiketoimintasääntöjä sovelletaan sovellusarkkitehtuurissa?</li> </ul>	<ul style="list-style-type: none"> <li>-Miten dataa ja arkistoja poistetaan järjestelmästä tai säilötään järjestelmään?</li> </ul>

(McGilvray 2008, s. 21)

## LIITE D: DATAN HALLINNAN ONGELMAT

	Suunnittelu	Kerääminen	Tallettaminen ja jakaminen
<b>Data (mitä)</b>	-Datan formaattia ja muotoilua ei ole määritelty ulkopuolelta tuleville toimituksille.	-Eri laboratoriot toimittavat datan eri muodossa.	-On kerätty paljon erityyppistä dataa, jota ei ole saatu tietokantaan. -Yksityiskohtia käytetystä analyysimenetelmästä ei saada linkitettyä tuloksiin.
<b>Prosessit (miten)</b>	-Datan viemistä tietokantaan ei ole otettu riittävän hyvin huomioon suunnitteluvaiheessa.	-Datan keräysmenetelmiä on monia, mikä tekee datan tuottamisen hallitsemisesta haastavaa. -Kenttämittausten prosessiin tulisi lisätä automaatiota tai tekoälyä. -Dataa toimitetaan kohdeyritykselle hyvin eri tavoilla.	
<b>Ihmiset ja organisaatiot (mitä)</b>	-Datan hallinnan ammattilainen tulee ottaa mittauksiin mukaan jo suunnitteluvaiheessa. -Datan tuottaminen on siiloista.	-Datan merkitys ei ole selvä kenttähenkilöstölle. -Perehdyttäminen datan tuottamiseen on haastavaa. -Kommunikointi tehtävistä/tehdyistä mittauksista on heikkoa.	-Roolitus ja vastuut eri datojen viemisestä eivät ole selkeät.
<b>Teknologia (miten)</b>		-Loggerien toiminta on epävarmaa. -Loggereista kerätään data eri menetelmillä. -Loggerien toimintaan tulisi lisätä automaatiota tai tekoälyä.	-Datalla on useita talletuspaikkoja. -Datan vieminen tietokantaan on hankalaa.

	Ylläpito	Käyttö
<b>Data (mitä)</b>	<ul style="list-style-type: none"> <li>-Datan valtava määrä on haaste ylläpidolle.</li> </ul>	<ul style="list-style-type: none"> <li>-Tuottodatan ja vanhojen tuulosraporttien saatavuus on huono.</li> <li>-Datan täydellisyydessä on puutteita.</li> <li>-Joidenkin taulukoiden johdonmukaisuudessa ja tulkittavuudessa on ongelmia.</li> <li>-Datan sijainti ei ole aina selvä.</li> </ul>
<b>Prosessit (miten)</b>	<ul style="list-style-type: none"> <li>-Virheellisen datan (esim. loggeridatan) manuaalinen korjaaminen on aikaa vievää ja työlästä.</li> <li>-Toiminnan laajentaminen palvelujen myymiseen asettaa haasteita datan ylläpidolle ja hallinnalle.</li> <li>-Virheiden korjaukselle ei ole selkeää prosessia tai protokollaa.</li> </ul>	<ul style="list-style-type: none"> <li>-Datan tarkastaminen / hyväksyminen on työlästä, ja siihen ei ole riittävästi aikaa.</li> <li>-Tarkastamatonta dataa käytetään raporteissa.</li> <li>-Datan käyttäminen voi tapahtua vasta vuosia datan keräämisen jälkeen.</li> </ul>
<b>Ihmiset ja organisaatiot (mitä)</b>	<ul style="list-style-type: none"> <li>-Datan hallinnan henkilöstö on pieni, mutta silti jokaisella on omat tarpeensa.</li> <li>-Resurssiongelma</li> </ul>	<ul style="list-style-type: none"> <li>- Ulkoisilla käyttäjillä on hydrologian dataa, jota sisäisillä ei ole.</li> <li>- Resurssiongelma aiheuttaa useita muita ongelmia.</li> </ul>
<b>Teknologia (miten)</b>	<ul style="list-style-type: none"> <li>-Datan muokkaaminen tietokannassa on vaikeaa.</li> <li>-Tietokannan kehittäminen on mahdotonta resurssiongelman takia.</li> <li>-Loggeridatan tietokantaan siirtämisestä vastuussa olevan tietojärjestelmän hallinta on hankalaa, sillä muutosten saaminen on toimittajariippuvaista.</li> </ul>	<ul style="list-style-type: none"> <li>-Tietokanta on vaikeakäyttöinen.</li> <li>-Työkalujen ja makrojen jatkuva kehitys on varmistettava, sillä tarpeiden muuttuessa työkaluja on kehitettävä.</li> </ul>